Statistische Datenanalyse mit SPSS 10 für Windows

Reihe Benutzereinführung
Band 25
2001

Universitäts-
Rechenzentrum

Universität Trier

Bernhard Baltes-Götz

Statistische Datenanalyse mit SPSS 10 für Windows

Benutzereinführung
Band 25
2001 (Rev. 0409)

Herausgeber:	Universitäts-Rechenzentrum Trier Universitätsring 15 D-54286 Trier WWW: http://www.uni-trier.de/urt/urthome.shtml E-Mail: urt@uni-trier.de Tel.: (0651) 201-3417, Fax.: (0651) 3921
Leiter:	Prof. Dr.-Ing. Manfred Paul
Autor:	Bernhard Baltes-Götz (E-Mail: baltes@uni-trier.de)
Druck:	Druckerei der Universität Trier
Copyright ©	2001; URT

Vorwort

SPSS (frühere Bedeutung: **S**tatistical **P**ackage for the **S**ocial **S**ciences, jetzige Interpretation: **S**uperior **P**erforming **S**oftware Systems) ist ein weitgehend komplettes und relativ leicht zu bedienendes Statistik-Programmpaket, das vor allem in den Wirtschafts- und Sozialwissenschaften sehr verbreitet ist und fast alle wichtigen Computertypen bzw. Betriebssysteme unterstützt (z.B. Windows, UNIX, Mac-OS).

Im vorliegenden Manuskript wird ein Einblick in die statistische Datenanalyse mit der SPSS-Version 10 vermittelt, wobei großer Wert gelegt wird auf die methodologische Einordnung der beschriebenen EDV-Techniken.
Zwar wird in erster Linie die *Windows*-Version behandelt, doch sind wesentliche Teile des Manuskripts wegen der weitgehend konsistenten Bedienungslogik auch für die SPSS-Versionen zu anderen Betriebssystemen verwendbar.

Das Manuskript wurde ursprünglich als Begleitlektüre zum URT-Kurs „Einführung in SPSS für Windows" erstellt, kann jedoch auch im Selbststudium verwendet werden. Dass dabei die meisten Themen in konkreter Arbeit am Rechner nachvollzogen werden sollten, folgt aus der Kurskonzeption:

Zielgruppe/Voraussetzungen

- Der Kurs ist konzipiert für Personen, die in wesentlichem Umfang bei Forschungsarbeiten mit SPSS mitwirken wollen, also z.B. im Rahmen einer Diplom- oder Dissertationsarbeit die Durchführung einer eigenen Studie planen oder bereits begonnen haben. Wer lediglich einfache Teilaufgaben zu erledigen hat (z.B. wenige Auswertungen mit einer bereits vorhandenen und fehlerbereinigten SPSS-Datendatei), der sollte eventuell zeitsparend anstatt des Kurses z.B. das SPSS-Online-Tutorium absolvieren.
- Im Kurs wird eine methodische Grundausbildung (empirische Forschung, Statistik) vorausgesetzt, wie sie üblicherweise in den Studiengängen empirisch orientierter Fächer vermittelt wird. Zwar werden im Kursverlauf viele methodische Themen in knapper Form behandelt, doch kann damit eher vorhandenes Wissen aufgefrischt als neues erworben werden.
- An EDV-Voraussetzungen werden elementare Fertigkeiten im Umgang mit PCs unter MS-Windows erwartet.

Kursinhalte

- In methodologischer Sicht kann der Kurs als partielles Repetitorium aufgefasst werden. In keinem Fall werden alle bei der empirischen Forschung potentiell relevanten Themen systematisch behandelt. Wir werden uns im wesentlichen darauf konzentrieren, wie mit dem **EDV-Werkzeug SPSS** die in anderen Veranstaltungen (z.B. zur empirischen Forschung oder Statistik) erlernten Methoden in der Praxis angewandt werden. Insbesondere kann die Anwendung der vielfältigen statistischen Auswertungsmethoden nur exemplarisch behandelt werden. Eine explizite Behandlung ist nur bei wenigen, besonders häufig eingesetzten Verfahren möglich (z.B. Kreuztabellenanalyse). Weil der Kurs in methodologischer Hinsicht als Repetitorium für *fortgeschrittene* Studierende gedacht ist, werden einfache Themen nicht explizit behandelt (z.B. Definition des arithmetischen Mittels).

Zu zahlreichen speziellen Auswertungsmethoden bietet das Rechenzentrum Spezialveranstaltungen an, in denen die wesentlichen methodologischen Grundlagen und natürlich die praktische Durchführung mit SPSS erläutert werden. Informationen über das URT-Kursprogramm finden Sie z.B. auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend über:

Rechenzentrum > Schulung/Kurse

Zu den meisten Kursen sind ausführliche Manuskripte entstanden, die Sie auf dem WWW-Server der Universität Trier von der Startseite ausgehend folgendermaßen erreichen:

Weitere Serviceangebote > DV-Dokumentationen > Elektronische Publikationen

- Im Sinne einer praxisnahen, projektorientierten Ausbildung beschreibt das vorliegende Manuskript eine vollständige empirische Studie von der ersten Idee über die Kodierung, Erfassung und Prüfung der Daten bis zur statistischen Auswertung. Dabei kommen die elementaren SPSS-Möglichkeiten zum Erfassen, Modifizieren und Analysieren von Daten zur Sprache, die bei den meisten Studien relevant sind. Wie bei jeder konkreten Studie sind darüber hinaus auch etliche spezielle Probleme aus verschiedenen Bereichen zu lösen (z.B. Untersuchungsplanung, Gestaltung der Erhebungsinstrumente, Transformation der erhobenen Variablen, Auswertungsfragen).
- Zwar werden auch in EDV-handwerklicher Sicht die SPSS-Möglichkeiten nicht annähernd vollständig behandelt, doch sollten Sie nach dem Kurs mit den erworbenen Grundkenntnissen unter Verwendung der aufgezeigten Informationsmöglichkeiten selbständig und erfolgreich mit SPSS arbeiten können.

Didaktische Hinweise

- Die oben beschriebene Projektorientierung erzwingt gelegentlich einen unkonventionellen Aufbau des Manuskriptes, weil alle Themen genau dann behandelt werden, wenn sie für die konkrete Beispiel-Studie relevant sind. Im Sinne einer guten Motivation aller Themen sind also geringfügige Einbußen bei der Systematik hinzunehmen.
- Im SPSS-Kurs am Rechenzentrum der Universität Trier sollen alle Teilnehmer(innen) ein kleines Forschungsprojekt selbständig durchführen (inkl. Datenerhebung und -erfassung). Als Begleitlektüre zu diesem Kurs enthält das vorliegende Manuskript entsprechende Anleitungen:

  Beschreibt ein Abschnitt eine Arbeitsabfolge, die nach Möglichkeit alle Leser konkret nachvollziehen sollten, dann steht am Anfang das Symbol , nötigenfalls ergänzt durch eine Beschreibung der Ausgangssituation. Ferner sollten die vorgeschlagenen Übungen durchgeführt werden. In den Übungen und in vielen Arbeitsabfolgen im Text wird die Beispiel-Studie allmählich weiterentwickelt. Wenn Entwicklungsschritte ausgelassen werden, fehlen eventuell Voraussetzungen für spätere Übungen.

  Gelegentlich fordert das Symbol  aber auch bei solchen Arbeitsabfolgen zum Mitmachen auf, die zwar für den Fortgang des Projektes weniger zentral sind, die aber wichtige Lernerfahrungen zu SPSS vermitteln und die durch Mitmachen erheblich leichter zu verstehen sind.

  Das Manuskript kann aber auch dann mit Gewinn gelesen werden, wenn auf die aktive Durchführung einer „eigenen“ Studie verzichtet wird. Für einen konkreten Datensatz werden alle Ergebnisse („Musterlösungen“) wiedergegeben.

Zugriff auf die Dateien zum Kurs

Leser(innen) im Selbststudium werden in der Regel keine eigene Datenerhebung durchführen, können jedoch anhand von Dateien, die auf Servern des Rechenzentrums zur Verfügung stehen, alle Projekt-Arbeitsschritte ab der Datenprüfung konkret durchführen. Im Internet finden Sie die Dateien ausgehend von der Startseite der Universität Trier (www.uni-trier.de) auf folgendem Weg:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen > Bedienungsanleitungen zu Statistikprogrammen > SPSS für Windows

Im Campusnetz der Universität Trier sind die Dateien noch bequemer über eine Netz-Freigabe zugänglich, nachdem Sie sich bei einem Windows-Rechner mit Einbindung in die NT-Domäe URT angemeldet haben. Führen Sie dort nach

Start > Ausführen

den Befehl

k baltes

aus, um die Netz-Freigabe als Laufwerk **K:** in Ihr Windows-System einzubinden.
Anschließend finden Sie die Dateien im Verzeichnis

K:\SPSS\Statistische Datenanalyse mit SPSS für Windows

Trier, im April 2001 Bernhard Baltes-Götz

Inhaltsverzeichnis

1 Von der Theorie zu den SPSS-Variablen

1.1 Statistik und EDV als Hilfsmittel der Forschung

Die Erfahrungswissenschaften bemühen sich um allgemeingültige Aussagen deskriptiver, explanatorischer oder prognostischer Art. In vielen Anwendungsbereichen sind dabei *deterministische* Gesetze kaum zu finden, und man muss sich auf die Untersuchung *probabilistischer* Gesetzte beschränken.

Beispiel: Welchen Effekt hat Nikotinkonsum auf die Entstehung von Lungenkrebs?
Wie wir wissen, führt dieselbe Nikotinbelastung bei verschiedenen Personen zu unterschiedlichen Folgen.

In einer solchen Situation können statistische Methoden dazu helfen, rationale Entscheidungen zu treffen, denn:

> "Statistics is a body of methods for making wise decisions in the face of uncertainty" (Wallis & Roberts, 1956, S. 1).

Die statistischen Methoden zur Entscheidungshilfe lassen sich in zwei Gruppen einteilen:

- **Deskriptive Statistik**
  Sie dient zur Darstellung und Zusammenfassung von *Stichprobendaten*. Hier kann man auch die *exploratorischen* Verfahren einordnen, deren Popularität in den letzten Jahren deutlich zugenommen hat.
- **Inferenzstatistik (schlußfolgernde Statistik)**
  Hier geht es darum, aus Stichprobendaten Informationen über die zugrunde liegende *Population* zu gewinnen. Die dabei verwendeten Methoden lassen sich wiederum in zwei Klassen einteilen:
  - **Schätzmethoden**
    Beispiel: Wie hoch ist bei Rauchern das Risiko, an Lungenkrebs zu erkranken? Hier ist eine Wahrscheinlichkeit zu schätzen.
  - **Hypothesentests** (konfirmatorische Verfahren)
    Beispiel: Ist bei Rauchern das Risiko für Lungenkrebs größer als bei Nichtrauchern?
    Hier ist eine Entscheidung zwischen zwei Hypothesen zu treffen:
    - *Nullhypothese*:
      Das Lungenkrebs-Risiko ist bei Rauchern nicht größer als bei Nichtrauchern.
    - *Alternativhypothese*:
      Das Lungenkrebs-Risiko ist bei Rauchern erhöht.

Die in den Beispielen zur Inferenzstatistik genannten Fragen sind anhand weniger, unrepräsentativer Einzelbeobachtungen (z.B. der steinalte Kettenraucher) nicht zu klären. Solche Anekdoten lassen keine sinnvollen Schlüsse und Entscheidungen zu, sondern demonstrieren lediglich die in obigem Zitat angesprochene Unsicherheit.

Eine grundlegende Strategie der statistisch arbeitenden Forschung, trotz Unsicherheit zu guten Entscheidungen zu kommen, besteht darin, zu einer Fragestellung möglichst viele, **unabhängige** Beobachtungen zu machen und diese mit einem statistischen Verfahren zufallskritisch zu analysieren. In obigen Beispielen wird man also bei möglichst vielen (z.B. 500) Personen (= **Beobachtungseinheiten, Merkmalsträgern, Fällen**) die **Merkmale** *Nikotinkonsum* und *Lungenkrebs-Erkrankungen* beobachten. Da außerdem eine Beteiligung weiterer Bedingungen an der

Lungenkrebs-Entstehung anzunehmen ist, wird man in einer wohldurchdachten Studie noch viele zusätzliche Merkmale erheben (z.B. Alter, Geschlecht, Beruf, Schadstoffbelastung des Wohnortes).

Eine praktikable Auswertung solcher Datenmengen ist aber nur mit EDV-Hilfe möglich. Mit **SPSS für Windows** steht ein bequemes, leistungsfähiges und sehr bewährtes Analysesystem für die statistische Forschung zur Verfügung. Es bietet fast alle wichtigen statistischen Verfahren sowie gute graphische Darstellungsmöglichkeiten und unterstützt alle in der Windows-Welt gebräuchlichen Verfahren zum Datenaustausch mit anderen Programmen (z.B. Zwischenablage, DDE, OLE, ODBC).

1.2 Planung und Durchführung einer empirischen Untersuchung im Überblick

Zunächst wollen wir uns einen Überblick über die verschiedenen Phasen eines empirischen Forschungsprojektes und damit auch über unser Kursprogramm verschaffen. Dabei werden zahlreiche Aufgaben, Methoden und Probleme angesprochen, über die Sie sich im Bedarfsfall in den Lehrveranstaltungen oder in der Literatur zur empirischen Forschung informieren können (siehe z.B. Bortz & Döring 1995, Pedhazur & Pedhazur Schmelkin 1991, Schnell, Hill & Esser 1989) .
Die anschließende Darstellung soll als Übersicht dienen und ist daher relativ knapp und abstrakt gehalten. Ihr folgt unmittelbar die konkrete und ausführliche Anwendung auf unsere Beispiel-Studie.

1.2.1 Forschungsziele bzw. -hypothesen

Einer empirischen Untersuchung wird in der Regel eine längere Phase der intensiven theoretischen Auseinandersetzung mit dem Thema vorangehen. Daraus ergeben sich Forschungsinteressen, die - u.a. in Abhängigkeit vom Forschungsstand - eher von explorativer (hypothesensuchender) oder eher von konfirmatorischer (hypothesenprüfender) Natur sind. In der Regel werden beide Forschungsstrategien vertreten sein. Die zu prüfenden Hypothesen müssen wegen ihrer Steuerungsfunktion für spätere Schritte möglichst exakt formuliert werden.

1.2.2 Untersuchungsplanung

Wenn Sie eine Theorie bzw. eine Hypothesenfamilie empirisch prüfen oder einen Gegenstandsbereich empirisch explorieren möchten, haben Sie bei der Untersuchungsplanung zahlreiche Aufgaben zu lösen:

- **Festlegung der Beobachtungseinheit und der zu untersuchenden Merkmale**
  In der Regel ergibt sich aus der Fragestellung unmittelbar, welche Beobachtungseinheiten (Merkmalsträger) Ihrer Studie zugrunde liegen sollten (z.B. Personen, Volkswirtschaften, Orte, Betriebe, Bodenproben), und welche Merkmale bei jeder Beobachtungseinheit festgestellt werden sollten.
  Beachten Sie unbedingt, dass die meisten statistischen Auswertungsverfahren **unabhängige** Beobachtungseinheiten voraussetzen. Wenn Sie etwa 50 Personen bitten, jeweils ein Brötchen mit Margarine und ein Brötchen mit Butter auf einer Skala von 1 (= "ungenießbar") bis 7 (= "ausgezeichnet") nach Geschmack zu bewerten, dann haben Sie **50** Beobachtungseinheiten, bei denen jeweils zwei Merkmale (Margarinebewertung und Butterbewertung) festgestellt werden. Es wäre falsch, hier von 100 Beobachtungseinheiten in zwei Gruppen auszugehen.
  Es soll noch ein weiteres Beispiel für eine Untersuchungsanlage angegeben werden, bei der die Beobachtungseinheit gelegentlich falsch festgelegt wird. Wenn in einer partnerschaftspsychologischen Untersuchung Ehepaare nach Ihrem Konfliktbewältigungs-

verhalten schriftlich befragt werden, wobei beide Partner denselben Fragebogen ausfüllen, dann ist nicht eine einzelne *Person* als Fall aufzufassen, sondern ein *Paar*.

- **Auswahl des Untersuchungsdesigns**
  Sie können z.B. einen (quasi-)experimentellen Untersuchungsplan wählen oder eine reine Beobachtungsstudie entwerfen, die z.B. quer- oder längsschnittlich angelegt sein kann.

- **Operationalisierung der zu untersuchenden Merkmale, Erstellung der Untersuchungsmaterialien**
  Sie werden bestrebt sein, objektive, reliable und valide Meßmethoden zu wählen bzw. zu entwerfen, die außerdem nicht zu aufwendig sind. Bei einer schriftlichen Befragung verdient z.B. die Formulierung der Fragen größte Aufmerksamkeit. Das Skalenniveau der gewählten Meßmethoden muss die Voraussetzungen der geplanten Auswertungsverfahren erfüllen.
  Bei den Überlegungen zur Operationalisierung von theoretischen Begriffen spielen auch die verfügbaren technischen Hilfsmittel für die Datenerhebung und -erfassung eine Rolle. Bei experimentellen Studien kommen oft Verfahren zur automatischen Messwerterfassung in Frage. Bei schriftlichen Befragungen großer Stichproben kann eventuell eine Anlage zum automatischen Einscannen und Interpretieren von Untersuchungsdokumenten rentabel eingesetzt werden. Voraussetzung ist dann u.a. die Beachtung einiger Regeln beim Entwurf der Untersuchungsmaterialien (siehe unten). Die computergestützte Diagnostik erlaubt eine interaktive und individualisierte Datenerfassung mit speziellen Vorzügen und Einschränkungen. Ähnliche Bedingungen sind mittlerweile in Online-Erhebungen (z.B. via Internet) zu realisieren, die nicht auf spezielle Rechner in einem Forschungslabor angewiesen sind.
  Am Rande möchte ich noch einen Fehler ansprechen, der überflüssigerweise recht häufig begangen wird: Wenn Sie das Glück haben, echte quantitative Variablen untersuchen zu können (z.B. Alter), dann sollten Sie den Informationsgehalt dieser Variablen *nicht* durch eine *künstliche* und *willkürliche* Klassenbildung reduzieren (z.B. durch Bildung der Altersklassen < 20, 21- 40, 41-60, > 60). Häufig sind Modelle für metrische Daten einfacher und erfolgreicher. Vor allem können Sie mit SPSS eine numerische Variable flexibel und *reversibel* in Klassen einteilen, wenn Sie dies für spezielle Analysen wünschen. Eine Ausnahme von dieser Regel ist vielleicht bei der Befragung von Personen nach ihrem Einkommen zu machen. Um bei dieser sensiblen Frage Widerstände zu vermeiden, muss man sich eventuell auf die Erhebung von groben Einkommensklassen beschränken.

- **Empirisch prüfbare Hypothesen**
  Aus einer in theoretischen Begriffen formulierten Hypothese ergibt sich im Verlauf der Untersuchungsplanung durch zahlreiche Konkretisierungen und Operationalisierungen eine in empirischen Begriffen formulierte und damit statistisch prüfbare Hypothese, die möglichst exakt notiert werden sollte. Dabei muss z.B. klar erkennbar sein, ob eine *gerichtete* oder eine *ungerichtete* Hypothese vorliegt.

- **Statistische Versuchsplanung**
  Für jede Hypothese ist ein statistisches Auswertungsverfahren zu wählen, dessen Voraussetzungen an Skalenniveau und Verteilungsverhalten der beteiligten Variablen (voraussichtlich) erfüllt sind.
  Zu jedem geplanten Test ist das Fehlerrisiko erster Art festzulegen, wobei z.B. die übliche 5%-Konvention übernommen werden kann.

Es ist zu überlegen, wie eine repräsentative und zur Durchführung der geplanten Auswertungsverfahren hinreichend große **Stichprobe** gewonnen werden kann. Bei ausgeprägt konfirmatorisch angelegten Studien sollte durch eine geeignete Wahl der Stichprobengröße der Fehler zweiter Art kontrolliert werden.[1]

- **Strukturierung und Kodierung der Daten**
  Am besten klärt man schon in der Planungsphase, wie die Merkmalsausprägungen kodiert werden sollen, damit sichergestellt ist, dass die erhobenen Informationen später auch sicher und bequem in die EDV übernommen werden können. Hierunter fällt z.B. die Festlegung, dass beim Merkmal Geschlecht die Ausprägung *weiblich* durch eine Eins und die Ausprägung *männlich* durch eine Zwei erfasst werden soll.
  Eigentlich sind also nur simple Konventionen festzulegen. Anfänger(innen) werden jedoch gerade bei der Konfrontation mit der „Maschine“, welche die forschungslogisch - und nicht etwa durch EDV-Restriktionen - vorgegebene Datenstruktur kompromisslos einfordert, konzeptionelle Probleme eher entdecken als etwa bei der schriftlichen Beschreibung ihres Forschungsvorhabens. Wir werden uns in Abschnitt 1.4 mit der Strukturierung und Kodierung von Daten ausführlich beschäftigen.
  Die Festlegungen zur Strukturierung und Kodierung der Projektdaten sollten in einem **Kodierplan** festgehalten werden, der später u.a. als genaue Arbeitsvorschrift für die Datenerfassung dient.

1.2.3 Durchführung der Studie (inklusive Datenerhebung)

Nach Abschluss der Planungs- und Vorbereitungsphase kann die Studie durchgeführt werden.

1.2.4 Datenerfassung und -prüfung

In der Regel liegen nach der Datenerhebung schriftliche Untersuchungsdokumente für jede Beobachtungseinheit vor. Damit fallen als nächstes folgende Arbeiten an:

- **Datenerfassung**
  Das Eintragen der Rohdaten in eine Datei auf der Festplatte eines Computers kann mit dem Dateneditor von SPSS geschehen, mit einem speziellen Datenerfassungsprogramm oder (fehleranfällig!) mit einem normalen Texteditor. In jedem Fall ist bei der Erfassung der in der Planungsphase oder spätestens nach der Datenerhebung definierte Kodierplan genau einzuhalten. Hier ist z.B. für jedes Merkmal festgelegt, wie seine Ausprägungen kodiert werden sollen (vgl. Abschnitt 1.4).
  Damit schriftliche Untersuchungsdokumente automatisch eingescannt werden können, müssen sie gewissen Designvorschriften genügen (siehe unten).
- **Überprüfung auf Erfassungsfehler**
  Je fehleranfälliger die gewählte Erfassungsmethode war, desto mehr Aufwand muss bei der Datenprüfung getrieben werden.

1.2.5 Datentransformation

Nach der Erfassung und Prüfung liegen bei vielen Studien die Daten immer noch nicht in auswertbarer Form vor. Vielfach müssen Variablen überarbeitet (z.B. rekodiert) oder aus Vorläufern neu berechnet werden (z.B. durch Mittelwertsbildung). Solche Transformationen machen nach einer Faustregel ca. **80%** aller SPSS-Arbeiten aus, die bei einer empirischen Studie nach Abschluss der Datenprüfung noch zu erledigen sind. Die eigentliche Daten*analyse*

[1] Bei der β-Fehler-basierten Kalkulation der Stichprobengröße kann z.B. das exzellente Programm Gpower eingesetzt werden. Eine Literaturangabe und eine kostenlose Bezugsquelle finden Sie in Abschnitt 8.

nimmt also mit 20% relativ wenig Arbeitszeit in Anspruch. Der Tendenz nach wird sich diese Aufteilung auch in unserem Kursprogramm widerspiegeln.
Natürlich können die obigen Prozentangaben in Abhängigkeit vom konkreten Projekt sehr stark variieren.

1.2.6 Statistische Datenanalyse

Nach langer Mühe können mit Hilfe von SPSS z.B. die gesuchten Schätzwerte ermittelt und die geplanten Hypothesentests durchgeführt werden. Bei einer eher explorativen Untersuchungsanlage ist eine längere, kreative Auseinandersetzung mit den Daten erforderlich, wobei zahlreiche Datentransformationen und statistische Analysen ausgeführt werden.

1.3 Beispiel für eine empirische Untersuchung [1]

Um die im Rahmen einer empirischen Untersuchung auftretenden EDV-Anwendungen unter realistischen Bedingungen üben zu können, wird im Verlauf des Kurses eine kleine psychologische Fragebogenstudie durchgeführt. Dabei werden Sie alle Phasen der empirischen Forschung von der ersten Idee bis zur statistischen Hypothesenprüfung mit Computerhilfe kennen lernen und die erforderliche Arbeiten zum großen Teil selbständig durchführen. Als Beispiel wurde u.a. deshalb eine psychologische Fragebogenstudie gewählt, weil die Kursteilnehmer dabei in wenigen Minuten interessante empirische Daten selbst erzeugen können. Damit ist auch die Phase der Daten*erhebung* in den Übungsablauf einbezogen, die ansonsten aus Zeitgründen ausgespart werden müsste.
Bezogen auf das in Abschnitt 1.2 vorgestellte Schema beschäftigen wir uns nun mit dem theoretischen Hintergrund unserer Studie und mit Fragen der Untersuchungsplanung.

1.3.1 Die allgemeinpsychologische KFA-Hypothese

Nach einer Theorie von Kahneman & Miller (1986) hängt die Stärke unserer emotionalen Reaktion auf ein positives oder negatives Ereignis u.a. davon ab, welche alternativen (aber nicht eingetretenen) Ereignisse wir uns vorstellen können, mit anderen Worten: welche **kontrafaktischen Alternativen** mental verfügbar sind. Wir wollen uns auf den Fall ungünstiger Ereignisse beschränken. Hierfür stellen Kahneman & Miller die folgende allgemeinpsychologische Hypothese auf:

> **Im Fall eines negativen Ereignisses erhöht die Verfügbarkeit kontrafaktischer (also positiver) Alternativen den erlebten Ärger.**

Im weiteren Verlauf wollen wir unser Projekt kurz als *KFA-Studie* bezeichnen.

1.3.2 Untersuchungsplanung

Hinsichtlich des Untersuchungsdesigns haben wir uns aufgrund praktischer Erwägungen bereits auf eine **querschnittlich angelegte Fragebogenstudie** festgelegt. Nun wollen wir die Beobachtungseinheit festlegen und die theoretischen Begriffe **operationalisieren** bzw. das Untersuchungsmaterial entwerfen.
Wir wollen die Untersuchungsteilnehmer bitten, sich in eine Geschichte einzufühlen, bei der zwei Personen objektiv denselben Schaden erleiden, jedoch in unterschiedlichem Grad eine kontrafaktische (also günstige) Alternative vor Augen haben. Dann sollen die Probanden für jeden Geschädigten angeben, wie stark sie sich in dessen Lage ärgern würden. Die genaue Instruktion ist dem unten wiedergegebenen Fragebogen (Teil 2) zu entnehmen.

---

[1] Hierbei werden in stark vereinfachter Form Ideen aus der Forschungsabteilung von Herrn Prof. Dr. J. Brandtstädter (Universität Trier) aufgegriffen, dem ich an dieser Stelle herzlich für die Erlaubnis und für die Überlassung von Untersuchungsmaterial danken möchte.

Die beiden Ärgermessungen werden durch Ratingskalen realisiert, wobei das Antwortformat der Anschaulichkeit halber an ein Thermometer mit den Ankerpunkten 0° und 100° erinnert. Wir gehen davon aus, dass die Ärgermessungen annähernd Intervallniveau besitzen.
Weil die Untersuchungsteilnehmer unabhängig voneinander agieren und Daten abliefern, spricht nichts dagegen, jede Person als **Beobachtungseinheit** aufzufassen.
In Abschnitt 1.3.1 wurde die KFA-Hypothese noch ohne Bezug auf unsere konkrete Untersuchung formuliert. Nun formulieren wir die beiden Hypothesen, zwischen denen wir uns später mit inferenzstatistischen Mitteln entscheiden wollen:

Nullhypothese: **Die Versuchspersonen erleben in der Rolle des Geschädigten mit hochgradig verfügbarer kontrafaktischer Alternative im Mittel *nicht* mehr Ärger als in der Rolle des Geschädigten mit "weit entfernter" kontrafaktischer Alternative.**

Alternativhypothese: **Die Versuchspersonen erleben in der Rolle des Geschädigten mit hochgradig verfügbarer kontrafaktischer Alternative im Mittel mehr Ärger.**

Wir wollen unser Entscheidungsproblem mit einem **t-Test für abhängige bzw. gepaarte Stichproben** lösen, falls die Verteilungsvoraussetzungen dieses Verfahrens erfüllt sind. Da gerichtete Hypothesen vorliegen, ist **einseitig** zu testen. Dabei wird eine Irrtumswahrscheinlichkeit erster Art in Höhe von **α = 5%** akzeptiert.

Unsere Studie soll aus praktischen Gründen mit der **studentischen Stichprobe** der Kursteilnehmer durchgeführt werden. Damit können unter induktivistischer Perspektive die Ergebnisse günstigstenfalls auf die Population der Studierenden generalisiert werden.

Da aus statistischer Sicht eine Stichprobe nie zu groß sein kann, sollen nach Möglichkeit alle Kursteilnehmer als Probanden gewonnen werden. Es ist aus praktischen Gründen nicht möglich, weitere Untersuchungsteilnehmer zu rekrutieren. Damit liegt der **Stichprobenumfang** fest, so dass wir auf eine β-Fehler-basierte Kalkulation verzichten.

Wie die mit unserem Fragebogen erfassten Merkmale in der EDV-Welt repräsentiert werden sollen, wird in Abschnitt 1.4 („Strukturierung und Kodierung der Daten“) behandelt.
Zuvor sollen noch einige zusätzliche Fragestellungen aufgegriffen und in den Untersuchungsplan aufgenommen werden.

1.3.3 Eine differentialpsychologische Hypothese

Neben der zentralen KFA-Hypothese soll in unserer Studie die folgende, auf Überlegungen von Scheier & Carver (1985) zurückgehende, differentialpsychologische Hypothese überprüft werden:

> **Der durch ein negatives Ereignis ausgelöste Ärger wird durch dispositionellen Optimismus gedämpft.**

Begründung: Dispositioneller Optimismus (im Sinne generalisierter positiver Ergebniserwartungen) führt zur Verwendung günstiger Bewältigungsstrategien (z.B. positive Reinterpretation, Akzeptieren der Realität).
Als Quasiereignis soll der schon zur Prüfung der allgemeinpsychologischen Hypothese verwendete imaginierte Schadensfall dienen (Fragebogenteil 2, s.u.).
Das arithmetische Mittel der für beide Situationsvarianten angegebenen Ärgerausprägungen soll uns als Ärgermaß dienen. Zur Erfassung von dispositionellem Optimismus wird der von Scheier & Carver (1985) entwickelte *Life Orientation Test* (LOT) eingesetzt (siehe Fragebogenteil 3).

Wie aus den Antworten auf die 12 Fragen dieses Tests ein Optimismus-Messwert zu ermitteln ist, wird später erläutert. Wir gehen jedenfalls davon aus, dass diese Messmethode annähernd Intervallniveau besitzt.
Nach dieser **Operationalisierung** der theoretischen Begriffe kann die folgende **empirisch prüfbare Alternativhypothese** formuliert werden:

> **Je höher der LOT-Wert einer Versuchsperson, desto weniger Ärger berichtet sie im Mittel für den imaginierte Schadensfall.**

Weil sich die Nullhypothese durch Negation der Alternativhypothese ergibt, muss sie nicht explizit notiert werden. Dementsprechend werden wir gelegentlich auch vom „Prüfen *der* differentialpsychologischen Hypothese“ sprechen, wenn wir die inferenzstatistisch begründete Entscheidung zwischen der Null- und der Alternativhypothese meinen.
Weil die Messungen zum Ärger und zum Optimismus (hoffentlich) auf Intervallskalenniveau vorliegen, kann die differentialpsychologische Hypothese mit einer **einfachen linearen Korrelationsanalyse** geprüft werden, sofern deren Modell- und Verteilungsvoraussetzungen erfüllt sind.
Die Hypothese ist wiederum **einseitig** formuliert und soll auf dem **5%-Niveau** geprüft werden.

1.3.4 Zum Einfluss demographischer Merkmale

Auf die Erfassung demographischer Merkmale (siehe Fragebogenteil 1) kann man in keiner Studie verzichten, auch wenn sich keine expliziten Hypothesen darauf beziehen. Man benötigt sie auf jeden Fall zur Beschreibung der Stichprobe, damit sich später die Leser(innen) von Berichten ein Urteil über die Interpretier- bzw. Generalisierbarkeit der Ergebnisse bilden können. Wir werden darüber hinaus einige demographische Merkmale auf Zusammenhänge mit unseren zentralen Projektvariablen untersuchen. Insofern finden sich auch in unserer überwiegend konfirmatorisch (hypothesenprüfend) angelegten Studie einige exploratorische Elemente.

1.3.5 Zu Übungszwecken miterhobene Merkmale

Ohne inhaltlichen Bezug, sondern nur zu Übungszwecken sollen zusätzlich folgende Informationen erhoben werden:

- Größe und Gewicht (siehe Fragebogenteil 1)
  Mit diesen Merkmalen lassen sich manche statistische Verfahren gut demonstrieren. Außerdem sorgen sie für das Auftreten gebrochener Zahlen in unseren Daten.
- Motive zur Kursteilnahme (siehe Fragebogenteil 4)
  Hier wollen wir die Behandlung von Mehrfachwahl-Fragen sowie von offenen Fragen üben.

1.3.6 Der Fragebogen

1) Angaben zur Person

Geschlecht	Frau ❒ Mann ❒
Geburtsjahr	
Fachbereich	
Körpergröße	__,__ __ m
Körpergewicht	__ __ kg

2) Fragen zur Reaktion in ärgerlichen Situationen

Versetzen Sie sich bitte möglichst gut in folgende Situation:

Herr Meier und Herr Schulze waren mit demselben Taxi auf dem Weg zum Flughafen. Sie sollten zur selben Zeit, aber mit verschiedenen Maschinen abfliegen. Durch einen Stau kommen sie erst eine halbe Stunde nach der planmäßigen Abflugzeit am Flughafen an.

***Herr Meier** erfährt, dass seine Maschine pünktlich vor einer halben Stunde gestartet ist.*

***Herr Schulze** erfährt, dass seine Maschine Verspätung hatte und erst vor zwei Minuten gestartet ist.*

Wie sehr würden Sie sich **ärgern**, wenn Sie in der Situation von ...

Herrn Meier wären?	0	10	20	30	40	50	60	70	80	90	100
Herrn Schulze wären?	0	10	20	30	40	50	60	70	80	90	100

Betrachten Sie bitte die Antwortskala als "Ärgerthermometer".

3) Aussagen zur Selbsteinschätzung

Teilen Sie bitte für die folgenden Selbstbeschreibungen durch Ankreuzen einer Antwortkategorie mit, inwiefern die Aussagen auf Sie persönlich zutreffen.

	völlig falsch	falsch	unentschiede	stimmt	stimmt genau
1. Auch in unsicheren Zeiten rechne ich im allgemeinen damit, dass sich alles zum besten wendet.	- -	-	0	+	+ +
2. Ich kann mich leicht entspannen.	- -	-	0	+	+ +
3. Wenn etwas schief gehen kann, dann passiert es mir auch.	- -	-	0	+	+ +
4. Bei allem sehe ich stets die negative Seite.	- -	-	0	+	+ +
5. Ich blicke kaum einmal mit Zuversicht in die Zukunft.	- -	-	0	+	+ +
6. Ich bin gern mit Freunden zusammen.	- -	-	0	+	+ +
7. Ich muss mich immer mit etwas beschäftigen.	- -	-	0	+	+ +
8. Ich habe stets die Hoffnung, dass die Dinge in meinem Sinne gehen.	- -	-	0	+	+ +
9. Die Dinge laufen immer so, wie ich es mir wünsche.	- -	-	0	+	+ +
10. Ich bin nicht leicht aus der Ruhe zu bringen.	- -	-	0	+	+ +
11. Ich glaube an den sprichwörtlichen "Silberstreifen am Horizont".	- -	-	0	+	+ +
12. Dass mir einmal etwas Gutes widerfährt, damit rechne ich kaum.	- -	-	0	+	+ +

4) Ihre Motive für die Teilnahme am SPSS-Kurs

a) Kreuzen Sie bitte in der folgenden Liste möglicher Motive für die Teilnahme am SPSS-Kurs alle für Sie zutreffenden Aussagen an:

Ich möchte SPSS kennen lernen, ...

❒ um eine eigene empirische Studie damit auszuwerten.
❒ weil in vielen Stellenanzeigen SPSS-Kenntnisse verlangt werden.
❒ weil ich mich um eine Stelle als EDV-Hilfskraft in der Forschung bewerben will (HIWI-Job).
❒ weil ich mich für EDV interessiere und ein modernes Programm kennen lernen möchte.
❒ weil ich mich für Statistik interessiere und mit Auswertungsverfahren experimentieren möchte.

❒ Für mich trifft keine der obigen Aussagen zu.

b) Möchten Sie im Kurs bestimmte statistische Methoden besonders gerne üben? Ja ❒ Nein ❒

Wenn "Ja", welche? ______________________________
____
____
____

1.4 Strukturierung und Kodierung der Daten

Die mit unserem Fragebogen zu erfassenden Merkmalsausprägungen müssen bei der späteren Datenerfassung in systematischer Form im Computer abgelegt werden. Dazu werden wir einen Kodierplan mit genauen Handlungsanweisungen für die Erfassung erstellen. Dabei müssen wir uns mit den Voraussetzungen beschäftigen, die SPSS für die Aufnahme unserer Daten bereitstellt. Diese sind in erster Linie durch die Logik der empirischen Forschung und nur in geringem Ausmaß durch EDV-Restriktionen festgelegt.
Die Kodierungs-Überlegungen sollten möglichst *vor* der Datenerhebung stattfinden, weil sich daraus Verbesserungen des Untersuchungsmaterials ergeben können.

1.4.1 Fälle und Merkmale in SPSS

Wir haben oben bereits daran erinnert, dass in einer empirischen Studie bei den einbezogenen **Fällen** bzw. **Beobachtungseinheiten** die Ausprägungen gewisser **Merkmale** festgestellt werden. Nun wollen wir uns ansehen, wie die Merkmalsausprägungen der Fälle im SPSS-System gespeichert werden. Die ganz konkrete Demonstration von KFA-Beispieldaten im SPSS-Dateneditorfenster wird das Verständnis der anschließenden, wieder eher allgemein-methodologisch geprägten, Ausführungen sicher unterstützen. U.a. werden dabei auch einige zentrale Begriffe des SPSS-Systems erläutert:

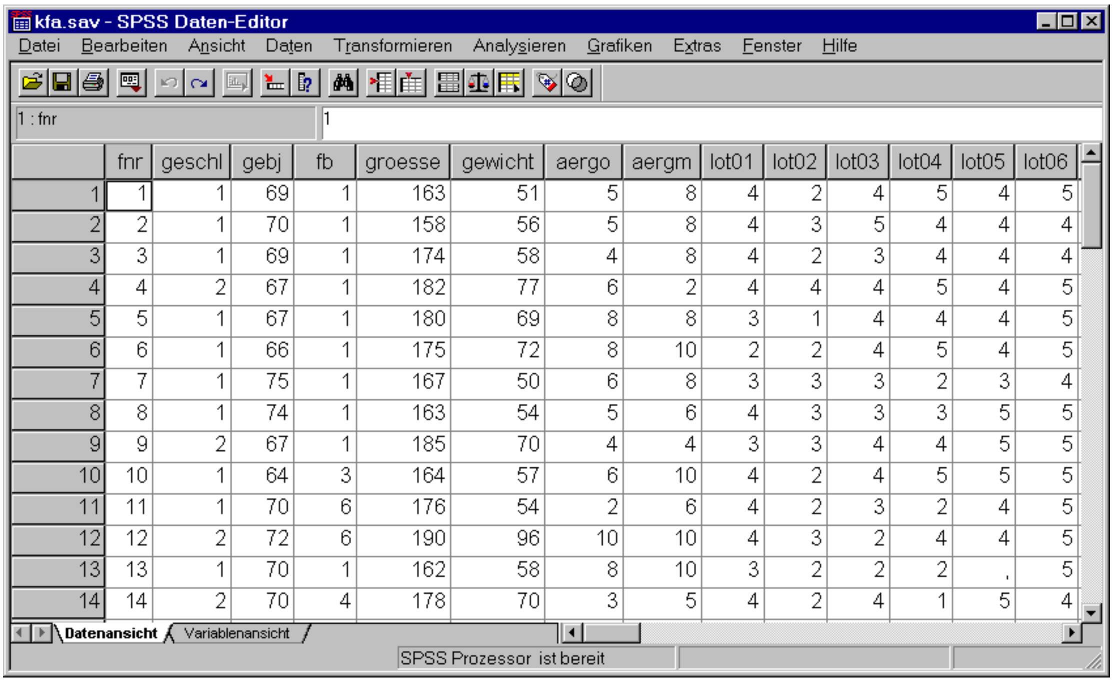
a) Variable

Der Begriff **Variable** wird in der Literatur zur statistischen Datenanalyse häufig synonym zu „Merkmal“ gebraucht. Wir wollen ihn SPSS-konform in einer etwas technischeren Bedeutung verwenden: Schreibt man für ein Merkmal die Ausprägungen aller Fälle in der Stichprobe untereinander, so entsteht ein Spaltenvektor. Genau einen solchen Spaltenvektor wollen wir als „Variable“ bezeichnen.

b) Datenmatrix und Dateneditor

Schreibt man die Spaltenvektoren aller Merkmale nebeneinander, so entsteht eine **Datenmatrix** (Datentabelle) mit den Variablen als Spalten und den Fällen als Zeilen.
Bei der Auswertungsarbeit mit SPSS ist diese Datenmatrix stets präsent: Sie kann bei der Datenerfassung im Fenster des **SPSS-Dateneditors** eingetragen und dort auch während der

laufenden Auswertung ständig eingesehen und bearbeitet werden. Die folgende Abbildung zeigt das Dateneditorfenster mit Beispieldaten zu unserem KFA-Projekt:



	fnr	geschl	gebj	fb	groesse	gewicht	aergo	aergm	lot01	lot02	lot03	lot04	lot05	lot06
1	1	1	69	1	163	51	5	8	4	2	4	5	4	5
2	2	1	70	1	158	56	5	8	4	3	5	4	4	4
3	3	1	69	1	174	58	4	8	4	2	3	4	4	4
4	4	2	67	1	182	77	6	2	4	4	4	5	4	5
5	5	1	67	1	180	69	8	8	3	1	4	4	4	5
6	6	1	66	1	175	72	8	10	2	2	4	5	4	5
7	7	1	75	1	167	50	6	8	3	3	3	2	3	4
8	8	1	74	1	163	54	5	6	4	3	3	3	5	5
9	9	2	67	1	185	70	4	4	3	3	4	4	5	5
10	10	1	64	3	164	57	6	10	4	2	4	5	5	5
11	11	1	70	6	176	54	2	6	4	2	3	2	4	5
12	12	2	72	6	190	96	10	10	4	3	2	4	4	5
13	13	1	70	1	162	58	8	10	3	2	2	2	.	5
14	14	2	70	4	178	70	3	5	4	2	4	1	5	4

Jede Variable, d.h. jede Spalte der (Fälle × Variablen)-Datenmatrix, ist durch ihren eindeutigen **Variablennamen** gekennzeichnet, über den sie bei der Anforderung statistischer oder graphischer Analysen angesprochen werden kann.

Nun wollen wir unsere eigenen Daten so strukturieren und kodieren, dass sie in eine Datenmatrix eingeordnet und mit SPSS verarbeitet werden können. Dabei soll u.a. angestrebt werden, den Aufwand und die Fehlergefahr beim Erfassen der Daten möglichst gering zu halten.

1.4.2 Strukturierung

Welche SPSS-Variablen im oben besprochenen Sinn sollen zur Aufnahme der mit unserem Fragebogen erfassten Informationen definiert werden? Obwohl die Antwort auf diese Frage trivial zu sein scheint, sind doch zu einigen Themen kurze Erläuterungen angebracht.

1.4.2.1 Variablen zur Fallidentifikation

Über die eigentlichen, empirischen, Variablen hinaus sollten in die Datenmatrix stets organisatorische Variablen aufgenommen werden, die eine Relation zwischen den schriftlichen oder sonstigen Untersuchungsdokumenten eines Falles und seinen Daten im Rechner herstellen. Eine solche Korrespondenz ist für eventuelle spätere Kontrollen oder Korrekturen der Daten unbedingt erforderlich. Meist verwendet man für diesen Zweck eine *einzelne* Variable, die z.B. FNR (für „Fallnummer") genannt werden kann. Natürlich muss die Fallidentifikation auch auf den schriftlichen oder sonstigen Untersuchungsdokumenten eingetragen werden.
Bei personbezogenen Daten wählt man aus Datenschutzgründen zur Fallidentifikation z.B. eine zufällig vergebene Nummer ohne jeden Bezug zu den Personalien.
Möglicherweise erscheint Ihnen das Eintippen einer Identifikations-Variablen sinnlos, weil im Dateneditor (siehe Abbildung in Abschnitt 1.4.1) die Zeilen bzw. Fälle ohnehin fortlaufend nummeriert sind. Die Nummern der Datenfensterzeilen stellen jedoch die gewünschte Korrespondenz zwischen den Datensätzen im Rechner und den nummerierten schriftlichen Untersuchungsunterlagen *nicht zuverlässig* her. Die SPSS-Numerierung der Datenfenster-Zeilen kann

sich nämlich leicht ändern, z.B. wenn ein Sortieren der Fälle nötig wird, oder wenn Fälle gelöscht oder eingefügt werden müssen.

1.4.2.2 Abgeleitete Variablen gehören nicht in den Kodierplan

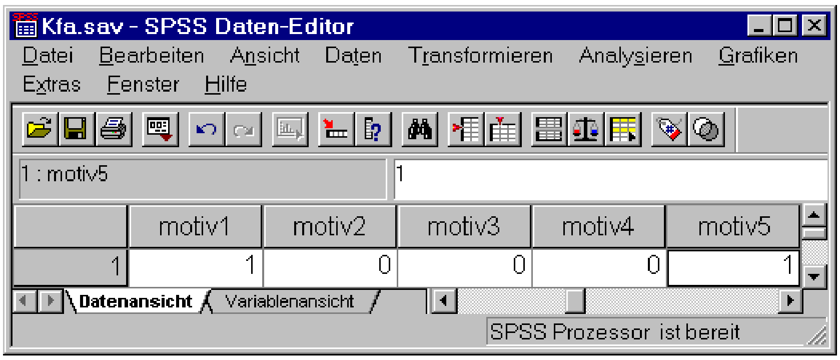
Häufig sind in einem Forschungsprojekt nicht nur die direkt erfassten Rohvariablen von Interesse, sondern auch darauf aufbauende Variablen. Im KFA-Projekt soll etwa der Optimismus der Untersuchungsteilnehmer(innen) durch ihre mittlere Antwort auf die LOT-Fragen geschätzt werden. SPSS verfügt über leistungsfähige Befehle zur Berechnung neuer Variablen aus bereits vorhandenen, so dass derartige Routinearbeiten keinesfalls während der Datenerfassung „per Hand“ ausgeführt werden sollten. Erfassen Sie also ausschließlich die Rohvariablen, und führen Sie alle erforderlichen Transformationen später mit SPSS-Methoden durch. Wir werden uns im weiteren Kursverlauf mit den SPSS-Transformationsmethoden ausführlich beschäftigen. Im Kodierplan mit den genauen Handlungsanweisungen für die Datenerfassung haben abgeleitete Variablen jedenfalls nichts zu suchen.

1.4.2.3 Mehrfachwahl-Fragen

Im Teil 4a unseres Fragebogens teilen die Untersuchungsteilnehmer *für jedes von fünf* möglichen Motiven mit, ob es bei ihrer Entscheidung für die Kursteilnahme relevant war. Damit erhalten wir von jeder Person *fünf* eigenständige Antworten und benötigen (ohne Komprimierungsverfahren, siehe unten) folglich in der SPSS-Datentabelle *fünf* Variablen, um die Antworten aufzunehmen, die wir z.B. durch die Zahlen 1, für „trifft zu“, und 0, für „trifft nicht zu“, kodieren können.
Beim Umgang mit einer solchen Mehrfachwahl-Frage müssen Sie sich vor allem vor dem aussichtslosen Versuch hüten, alle Antworten auf die Frage in *eine* Variable zu verpacken. Dies käme dem unsinnigen Versuch gleich, *mehrere* Werte (z.B. Zahlen) in *eine* Zelle der SPSS-Datenmatrix einzutragen.

1.4.2.3.1 Vollständige Sets aus dichotomen Variablen

In unserem Beispiel führt also eine Mehrfachwahl-Frage zu fünf dichotomen SPSS-Variablen, die jeweils die Information darüber enthalten, ob ein bestimmtes Motiv vorlag oder nicht.
Das folgende Datenfenster zeigt die fünf Variablen, hier bezeichnet mit MOTIV1 bis MOTIV5, bei einem Fall mit dem Antwortmuster „1,0,0,0,1“:



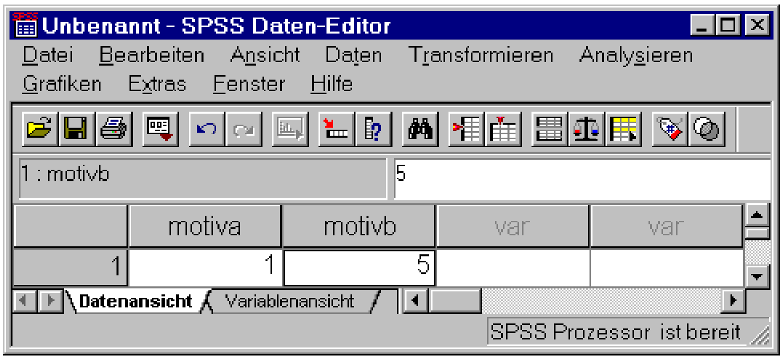
Wir werden später ein sogenanntes **Mehrfachantworten-Set** bestehend aus diesen fünf Variablen definieren und mit dessen Hilfe eine gemeinsame Auswertung der Variablen vornehmen. An dieser Stelle müssen Sie jedoch unbedingt akzeptieren, dass wir es mit *fünf* Merkmalen bzw. Variablen zu tun haben.

1.4.2.3.2 Sparsame Sets aus kategorialen Variablen

Das im letzten Abschnitt beschriebene Schema zur Übersetzung einer Mehrfachwahl-Frage in mehrere SPSS-Variablen ist angemessen, sofern nicht zu viele Antwortmöglichkeiten vorgegeben werden. Wenn Sie etwa eine Liste mit 100 möglichen Freizeitaktivitäten präsentieren, dann führt das Schema zur Definition von 100 SPSS-Variablen. Unter der Annahme, dass jeder einzelne Untersuchungsteilnehmer maximal sieben verschiedene Optionen ankreuzen wird, ist das Schema sicherlich unökonomisch und unpraktisch. Für solche Situationen bietet sich ein alternatives Vorgehen an, das im eben konstruierten Freizeit-Beispiel lediglich sieben Variablen bzw. Spalten in der SPSS-Datentabelle benötigt. Auch dieses „Komprimierungsverfahren“ soll an unserem Motiv-Beispiel demonstriert werden, obwohl es in diesem Fall (bei nur fünf Antwortmöglichkeiten) sicher weniger geeignet ist. Unter der Annahme, dass pro Person maximal *zwei* verschiedene Motive zutreffen werden, definieren wir die beiden SPSS-Variablen MOTIVA und MOTIVB, die jeweils folgende Werte annehmen sollen:

1 für das Motiv „Eigene empirische Studie“,
2 für das Motiv „Orientierung am Arbeitsmarkt“,
3 für das Motiv „Bewerbung als EDV-Hilfskraft“,
4 für das Motiv „Interesse an der EDV“,
5 für das Motiv „Interesse an Statistik“

Mit den Variablen MOTIVA und MOTIVB stehen für jede Person zwei Möglichkeiten zur Verfügung, um die „Hausnummern“ von angekreuzten Motiven einzutragen. Das Antwortmuster „1,0,0,0,1“ wird folgendermaßen übertragen:



Im Prinzip kann man im Beispiel die beiden Werte Eins und Fünf auch in umgekehrter Reihenfolge eintragen (MOTIVA = 5, MOTIVB = 1). Wesentlich ist nur, dass die Nummer jedes angekreuzten Motivs in einer der beiden Variablen als Wert auftritt. Von einer Person, die zwei Motive angekreuzt hat, wissen wir *nicht*, welchem Motiv sie die größte Bedeutung beimisst. Daher können auch die resultierenden Variablen eine solche subjektive Ranginformation nicht enthalten. Allerdings wird man beim Erfassen der Systematik halber wohl so vorgehen, dass in MOTIVA die Nummer des ersten angekreuzten Motivs (bei Anordnung von oben nach unten) eingetragen wird usw.

Wir sparen drei Variablen ein, wobei kein Informationsverlust eintritt, *wenn* tatsächlich pro Person maximal zwei Motive angekreuzt werden.

Auch bei der sparsamen Informationsanordnung kann man mit SPSS z.B. für jedes Motiv ermitteln, wie viel Prozent der Kursteilnehmer es angekreuzt haben. Vor einer solchen Auswertung ist wiederum ein Mehrfachantworten-Set zu definieren, diesmal bestehend aus den beiden Variablen MOTIVA und MOTIVB, wobei in der zugehörigen SPSS-Dialogbox eine *kategoriale* Kodierung der Variablen anzugeben ist. Dazu später mehr.

Bei manchen Auswertungen erfordert die sparsamere Art der Informationsanordnung einen zusätzlichen Transformationsschritt, so dass sie nur bei deutlichem Einsparungseffekt verwendet werden sollte.

1.4.2.4 Offene Fragen

Offene Fragen lösen vielfältige und oft schwer strukturierbare Antworten aus, und es bleibt dann offen, ob und wie die Antworten in SPSS-Variablen übersetzt werden sollen. Im Fall unseres Fragebogenteils 4b könnten wir z.B. nach grober Sichtung der Antworten ein Kategorienschema für statistische Auswertungsverfahren entwickeln. Bei der Umsetzung in SPSS-Variablen können wir dann genauso vorgehen, wie es im letzten Abschnitt (über Mehrfachwahl-Fragen) beschrieben wurde. Generell wird man bei einem relativ kleinen Kategorienschema ein vollständiges Set mit dichotomen Variablen verwenden, ansonsten ein sparsames Set aus kategorialen Variablen (siehe oben).
Aus unserer Frage 4b nach speziellen methodischen Interessen könnte so z.B. ein vollständiges Set mit den folgenden dichotomen Variablen entstehen:

REG	für die Regressionsanalyse
KT	für die Kreuztabellenanalyse
FAKT	für die Faktorenanalyse

Bei der Variablen REG wäre eine Eins einzutragen, wenn ein Fall auf die offene Frage hin die Regressionsanalyse angegeben und damit sein Interesse an dieser Methode signalisiert hätte. Anderenfalls müsste eine 0 notiert werden. Diese 0 wäre aber *nicht* als explizit bekundetes Desinteresse an der Regressionsanalyse zu interpretieren.
Die oben vorgeschlagene grobe Sichtung der Antworten erfordert eventuell eine recht aufwendige „manuelle" Vorauswertung der Fragebögen, die sich mit folgendem Trick vermeiden lässt: Man verwendet ein *dynamisches* Kategorienschema in Verbindung mit einem sparsamen Set kategorialer Variablen. In unserem Beispiel kann man z.B. mit einem sparsamen Set aus fünf Variablen für jeden Fall bis zu fünf spezielle Auswertungsinteressen festhalten. Das Kategorienschema wird erst während der Datenerfassung entwickelt, indem man bei jedem Fall entscheidet, in welche bereits definierten oder neu aufzunehmenden Kategorien seine Antworten einzuordnen sind. Das Schema kann dynamisch um beliebig viele Kategorien erweitert werden, da die fünf Variablen beliebig viele verschiedene Werte als Kategoriennummern aufnehmen können. Selbstverständlich müssen die neu aufgenommenen Kategorien mit den vergebenen Nummern sorgfältig dokumentiert werden. Falls mehrere Personen an der Erfassung beteiligt sind, muss die eindeutige Zuordnung durch entsprechende Verabredungen sichergestellt werden.

Offene Fragen sind sicher vielfach sinnvoll, weil sie Informationen zutage fördern können, an die bei der Untersuchungsplanung niemand gedacht hat. Gelegentlich ist jedoch die Strukturierung der Antworten so aufwendig und problematisch, dass eine *statistische* Analyse nicht sinnvoll erscheint.

1.4.3 Kodierung

Für jedes erhobene Merkmal muss festgelegt werden, wie die einzelnen Merkmalsausprägungen kodiert werden sollen. Dabei ist eine Kodierung durch einfach aufgebaute Werte anzustreben (z.B. durch positive, ganze Zahlen). Bei konkreten Überlegungen zur Kodierung müssen wir berücksichtigen, welche Variablentypen von SPSS unterstützt werden:

1.4.3.1 Die wichtigsten Variablentypen in SPSS

An dieser Stelle beschränken wir uns auf die wichtigsten Variablentypen, mit denen die meisten Projekte auskommen:

- **Numerische Variablen**
  Werte: reelle Zahlen
  Z.B. geeignet für die Merkmale:
  - Größe
  - Gewicht

- **String-Variablen (synonym: alphanumerische Variablen, Zeichenkettenvariablen)**
  Werte: Folgen von Zeichen (Buchstaben, Ziffern, Sonderzeichen), maximale Länge: 255
  Z.B. geeignet für die Merkmale:
  - Familienname
  - Man könnte das Merkmal *Geschlecht* alphanumerisch kodieren mit den Werten *weiblich* und *männlich*.

- **Datumsvariablen**
  Werte: Datumsangaben
  Z.B. geeignet für das Merkmal: Geburtsdatum

Anwendungsfälle für Datumsvariablen, die in unserer KFA-Studie übrigens nicht benötigt werden, dürften in der Regel klar erkennbar sein. Ansonsten müssen Sie sich also nur zwischen der numerischen und der alphanumerischen Kodierung entscheiden.
Bei Merkmalen mit **mindestens ordinalem Skalenniveau** ist offensichtlich nur die numerische Kodierung sinnvoll.
Bei Merkmalen mit **Nominalskalenniveau** hat man hingegen die Wahl zwischen numerischer und alphanumerischer Kodierung der Merkmalsausprägungen.

Beispiel Geschlecht:
- numerische Kodierung: **1** für Frauen, **2** für Männer
- alphanumerische Kodierung: **f** für Frauen, **m** für Männer

Beim Arbeiten mit SPSS empfiehlt es sich, auch nominalskalierte Merkmale numerisch zu kodieren, weil manche Auswertungsverfahren für diese Merkmale nur numerische Variablen akzeptieren (z.B. die Diskriminanzanalyse).[1]

1.4.3.2 Das Problem fehlender Werte

Trotz aller Sorgfalt sind in fast jedem Forschungsprojekt bei manchen Fällen einige Variablenausprägungen nicht bekannt, z.B. wegen technischer Fehler oder wegen nachlässig ausgefüllter Fragebögen. Bei der Kodierungsplanung muss daher festgelegt werden, was im Falle fehlender Werte in die zugehörige Zelle des Datenfensters eingetragen werden soll. Diese Ersatzwerte bezeichnet man häufig als "MD-Indikatoren", wobei "MD" für "missing data" steht. Gelegentlich sind bei einer Variablen sogar mehrere MD-Indikatoren nötig, wobei z.B. ein erster Indikator signalisiert "Frage trifft nicht zu" und ein zweiter bedeutet "Keine auswertbare Antwort vorhanden".

Beispiel: Angenommen, wir hätten uns im demographischen Teil unseres Fragebogens danach erkundigt, ob ein Teilnehmer Wehr- bzw. Zivildienst abgeleistet hat (0 = Nein, 1 = Ja). Dann könnten wir zu dieser Frage die SPSS-Variable DIENST definieren und dabei u.a. folgende Kodierungsregeln vereinbaren:
  - Frauen erhalten bei DIENST den Wert 8 (≘ "Frage trifft nicht zu").
  - Verweigert ein Mann die Antwort, erhält er den Wert 9.

Beachten Sie bei der Verwendung von benutzerdefinierten MD-Indikatoren folgende Regeln:

- Es ist klar, dass alle MD-Indikatoren einer Variablen außerhalb des validen Wertebereichs liegen müssen. So wäre z.B. die 99 kein geeigneter MD-Indikator für unsere Variable Körpergewicht (gemessen in kg).

[1] Offenbar überarbeitet SPSS sukzessive alle Prozeduren dahingehend, dass auch *kurze* String-Variablen (mit maximal achtstelligen Werten) akzeptiert werden, wenn in statistischer Hinsicht nur Nominalskalenniveau erforderlich ist. Diese Anpassung ist jedoch noch nicht für alle Prozeduren erfolgt.

- Wählen Sie möglichst prägnante oder extreme Werte (also z.B. bei einer Variablen mit den validen Werten 1 und 2 den MD-Indikator 9). Dies bewirkt warnend auffällige Ergebnisse, falls Fälle mit fehlenden Werten nicht ordnungsgemäß von einer Analyse ausgeschlossen werden.
- Der Einfachheit halber sollte für alle Variablen mit ähnlichem Wertebereich derselbe MD-Indikator verwendet werden. Sie können aber selbstverständlich für jede Variable individuelle MD-Indikatoren festlegen.

Wichtig: **Für jede betroffene Variable müssen dem SPSS-System alle benutzerdefinierten MD-Indikatoren bekannt gemacht werden.**

1.4.3.2.1 System-Missing (SYSMIS)

Neben den vom Benutzer variablenspezifisch vereinbarten MD-Indikatoren verwendet SPSS für *alle numerischen* Variablen automatisch einen weiteren MD-Indikator, der mit „System-Missing“, „systemdefiniert fehlend“ oder „SYSMIS“ bezeichnet wird. Er wird von SPSS als Wert eines Falles für eine Variable immer dann automatisch benutzt, wenn eines der folgenden Probleme auftritt:

- Im Dateneditor (siehe unten) bzw. beim Lesen einer bereits vorhandenen Datendatei findet SPSS im Feld einer als numerisch definierten Variablen unzulässige Zeichen oder überhaupt keinen Eintrag.
- Beim Neuberechnen einer Variablen aufgrund einer Transformationsanweisung (siehe unten) fehlt ein Argument, oder der Funktionswert ist nicht definiert (z.B. bei Division durch 0).

Wir haben gerade erfahren, dass SPSS leere Eingaben im Dateneditor selbsttätig durch den automatisch definierten MD-Indikator System-Missing ersetzt. Wenn wir diesen Ersatzwert eintragen wollen, müssen wir also lediglich die betroffene Zelle leer lassen.

Tipp: Bei der Datenerfassung mit dem SPSS-Dateneditor können Sie routinemäßig als MD-Indikator den Wert System-Missing verwenden, eventuell ergänzt durch zusätzliche benutzerdefinierte MD-Indikatoren. System-Missing kann bequem dadurch eingetragen werden, dass die betroffene Zelle im Datenfenster einfach leer gelassen wird. Weil SPSS den Wert System-Missing automatisch richtig versteht, ist eine Deklaration nicht nötig und kann daher auch nicht vergessen werden.

Da in jeder numerischen Zelle des Dateneditors eine reelle Zahl zu stehen hat, ist auch SYSMIS letztlich nur eine normale Zahl, aber eine, die als valider Messwert (hoffentlich!) nie vorkommt: SYSMIS = $-1{,}7976931 \cdot 10^{308}$. Im Datenfenster und in der SPSS-Ergebnisausgabe wird SYSMIS durch das länderspezifische Dezimaltrennzeichen dargestellt, bei uns also durch ein Komma (siehe Abbildung in Abschnitt 1.4.1, Variable LOT05 bei Fall 13).

1.4.3.2.2 Fehlende Werte bei Mehrfachwahl-Fragen und offenen Fragen

Nachdem der Sinn und die Verwendung von MD-Indikatoren geklärt sind, geht es in diesem Abschnitt um eine spezielle Interpretationsunsicherheit im Zusammenhang mit fehlenden Werten, die bei Mehrfachwahl-Fragen aus der Verwendung eines Probanden-freundlichen Antwortformates resultiert:
Vielleicht haben Sie sich schon Gedanken darüber gemacht, wozu im Fragebogenteil 4a zu den Motiven für die Kursteilnahme die Ankreuzalternative „Für mich trifft keine der obigen Aussagen zu“ nützlich sein könnte. Ohne diese Möglichkeit, die Abwesenheit der fünf vorgegebenen Motive explizit zu bekunden, könnten wir bei einem Fragebogen mit fünf leeren Motivkästchen folgende Möglichkeiten nicht unterscheiden:

- Bei der Person trifft tatsächlich keines der Motive zu.
- Die Person hat den Fragebogenteil 4a nicht bearbeitet (fehlende Daten).

Ursache für die Interpretationsunsicherheit ist offenbar das vereinfachte Antwortformat, das pro Motiv nur *ein* Kästchen vorsieht, statt jeweils ein Ja- *und* ein Nein-Kästchen vorzugeben. Damit ersparen wir den Untersuchungsteilnehmern zahlreiche Nein-Markierungen. Dies ist sinnvoll, damit ihre Motivation nicht überstrapaziert wird, und die Fehlerquote gering bleibt.

Bei der offenen Frage in Teil 4b wird durch die vorgeschaltete Frage, ob überhaupt bestimmte Methoden gewünscht sind, dafür gesorgt, dass bei Fragebögen ohne eingetragene Methoden-interessen folgende Möglichkeiten unterschieden werden können:

- Die Person hat kein Interesse an speziellen Auswertungsmethoden.
- Die Person hat den Fragebogenteil 4b nicht bearbeitet (fehlende Daten).

1.4.3.2.3 Auch komplizierte Überlegungen können das Leben erleichtern

Im Zusammenhang mit dem MD-Problem bei den Variablen zu unserem Fragebogenteil 4 wage ich nun einige Vorschläge, die zwar dem Datenerfasser das Lebern erleichtern, aber zuge-gebenermaßen die Kursteilnehmer(innen) beim ersten Entwurf eines Kodierplans durch einige zusätzliche Überlegungen belasten:
Bei der Mehrfachwahl-Frage nach den Kursmotiven haben wir geschickt durch die sechste Ankreuzalternative „Für mich trifft keine der obigen Aussagen zu“ dafür gesorgt, dass wir Personen mit fehlenden Werte sicher identifizieren können. Wir könnten den Erfasser nun im Kodierplan beauftragen:

> "Schreibe bei den Variablen MOTIV1 bis MOTIV5 den Wert 1, wenn das zugehörige Käst-chen markiert ist, sonst eine 0.
> Wenn aber weder die sechste Antwortalternative angekreuzt, noch ein spezielles Motiv ge-wählt ist, dann versorge die Variablen MOTIV1 bis MOTIV5 mit einem MD-Indikator."

Statt dessen wollen wir die Reaktion der Teilnehmer auf die sechste Antwortalternative in formaler Analogie zum Vorgehen bei den anderen Antwortmöglichkeiten durch eine eigene SPSS-Variable erfassen, die KEINE genannt werden soll. Das in Abschnitt 1.4.2.3.1 gezeigte Datenfenster wird nach diesem Vorschlag folgendermaßen erweitert:



Der doppelt motivierte Beispielfall hat die letzte Alternative korrekterweise nicht angekreuzt, so dass er bei KEINE den Wert 0 erhält.
Dem Erfasser können wir nun eine einfachere Regel für die Variablen MOTIV1 bis KEINE angeben:

> "Schreibe bei *markierten* Kästchen eine 1, sonst eine 0."

Damit wird die Lösung des MD-Problems zugunsten einer möglichst einfachen Datenerfassung in die spätere Projektphase der Datentransformation verschoben, wobei uns mächtige Hilfsmittel des SPSS-Systems zur Verfügung stehen werden. Die unter der eben angegebenen Regel

erfassten Variablen zum Fragebogenteil 4a (MOTIV1 bis KEINE) werden wir mit SPSS-Transformationsanweisungen so behandeln, dass die Ausprägungen zuverlässig folgendermaßen interpretiert werden können:

0	=	nein
1	=	ja
System-Missing	=	Wert unbekannt

Um dies zu erreichen, müssen wir uns mit folgenden Konstellationen beschäftigen:

- Hat ein Fall bei den Variablen MOTIV1 bis MOTIV5 **und** bei der Variablen KEINE den Wert 0, gehen wir von einem MD-Problem aus und setzen alle genannten Variablen auf den Wert SYSMIS.
- Wenn eine Person mindestens ein vorgegebenes Motiv angekreuzt hat, aber gleichzeitig auch die Alternative "Für mich trifft keine der obigen Aussagen zu", dann gehen wir von einem harmlosen Fehler aus und setzen die Variable KEINE auf 0.

Zur Erfassung der Informationen im Fragebogenteil 4b wollen wir ein dynamisches Kategorienschema mit einem zugehörigem sparsamen Set kategorialer Variablen METH1 bis METH5 (vgl. Abschnitt 1.4.2.4) entwickeln. Der damit schon reichlich belastete Erfasser soll folgendermaßen vorgehen:

- Die Antwort auf die Frage, ob spezielle Methodenwünsche bestehen, wird konventionell in der Variablen SMG mit folgender Kodierungsvorschrift erfasst:

  Kreuz bei „Nein“: Schreibe eine 0
  Kreuz bei „Ja“: Schreibe eine 1
  Keine Antwort: Lasse die Zelle leer (= SYSMIS)

- In die Dateneditor-Zellen zu den Variablen METH1 bis METH5 sollen die Kategoriennummern der gewünschten Methoden eingetragen werden. Bei weniger als 5 Nennungen, sollen die nicht benötigten Variablen einfach leer bleiben.

Die Variablen SMG sowie METH1 bis METH5 müssen später mit SPSS-Transformationsanweisungen so behandelt werden, dass ihre Ausprägungen zuverlässig folgendermaßen interpretiert werden können:

- Bei SMG:

0	=	nein
1	=	ja
System-Missing	=	Wert unbekannt

- Bei METH1 bis METH5:

0	=	Von der *i*-ten (*i* = 1,..5) Option zur Nennung einer interessierenden Methode wurde kein Gebrauch gemacht.
natürliche Zahl ≥ 1	=	Die Methode mit dieser Kategoriennummer wurde angegeben.
System-Missing	=	Wert unbekannt

Dazu müssen unter den verschiedenen Wertekonstellationen folgende Anpassungen vorgenommen werden:

		Mindestens eine speziell interessierende Methode angegeben?	
		Ja	Nein
SMG	1	METH1 ... METH5: SYMIS → 0 Bem.: Korrektes Antworteverhalten. Variablen zu nicht benutzten Optionen (gem. Kodierplan bisher auf SYSMIS) werden auf 0 umgesetzt.	SMG: 1 → SYMIS Bem.: Irreguläres Antwortever-halten. METH1 bis METH5 behalten SYMIS. SMG wird ebenfalls auf SYMIS gesetzt.
	0	SMG: 0 → 1 METH1 ... METH5: SYMIS → 0 Bem.: „Leicht" irreguläres Antworteverhalten. Wir sind großzügig und passen SMG an.	METH1 ... METH5: SYMIS → 0 Bem.: Korrektes Antworteverhalten. Die Variablen zu allen Optionen (gem. Kodierplan bisher auf SYMIS) werden auf 0 umgesetzt.
	SYSMIS	SMG: SYSMIS → 1 METH1 ... METH5: SYMIS → 0 Bem.: „Leicht" irreguläres Antworteverhalten. Wir sind großzügig und passen SMG an.	Bem.: Irreguläres Antworteverhalten. Alle Variablen behalten den Wert SYMIS.

Ich bin mir darüber im Klaren, dass beim Lesen der letzten Ausführungen wenig Freude aufkam. Im Sinne einer guten Vorbereitung auf die Praxis waren sie allerdings vertretbar. Eine exakte Lösung des MD-Problems ist leider oft mühselig und zeitaufwendig. In Abschnitt 1.2 war davon die Rede, dass Datentransformationen ca. **80%** aller SPSS-Arbeiten ausmachen, die bei einer typischen empirischen Studie nach Abschluss der Datenprüfung noch zu erledigen sind. Von diesen 80% ist ein erheblicher Anteil durch das Problem der fehlenden Werte bedingt.
Obige Tabellen werden uns außerdem im Abschnitt über bedingte Datentransformationen vorzügliche Übungsmöglichkeiten bieten und sind vielleicht schon dadurch gerechtfertigt.

1.4.3.3 Fehlerquellen bei der manuellen Datenerfassung minimieren

Wenn die Daten manuell erfasst werden (siehe unten), ist bei den Kodierungsvereinbarungen darauf zu achten, dass dem Erfasser keine zeitaufwendigen und fehleranfälligen Arbeiten zugemutet werden, z.B.:

- Treten gebrochene Zahlen als Werte auf (z.B. bei unserer Frage nach der Körpergröße), so kann man durch Wechsel der Maßeinheit das lästige Dezimalkomma eliminieren.
  Beispiel: 1,65 m → 165 cm
- Bei bipolaren Skalen mit positiven und negativen Werten (z.B. bei unseren LOT-Fragen) empfiehlt sich eine Transformation zu ausschließlich positiven Werten z.B.:

- -	→	1
-	→	2
0	→	3
+	→	4
+ +	→	5

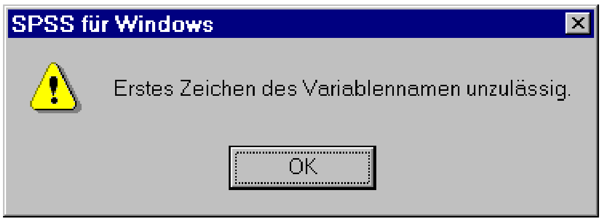
  Vorteil: Im Vergleich zu einer „bipolaren" Kodierung von –2 bis +2 spart man Tipparbeit und macht keine Fehler durch vergessene Vorzeichen bei den negativen Zahlen.
- Wurden einige Fragen aus meßtechnischen Gründen umgepolt (negativ formuliert), was im KFA-Projekt bei einigen LOT-Fragen geschehen ist, so sollte diese Umpolung *keinesfalls* während der Erfassung rückgängig gemacht werden. Dies geht sehr viel bequemer und ohne Fehlerrisiko mit den Transformationsmöglichkeiten von SPSS (siehe unten).

1.4.3.4 SPSS-Variablennamen

Es empfiehlt sich, an dieser Stelle auch schon SPSS-Namen für die Variablen festzulegen und ebenfalls in den Kodierplan (siehe Abschnitt 1.4.3.5) aufzunehmen. Dabei sind die SPSS-Regeln für Variablennamen zu beachten:

- Maximal acht Zeichen
- Das erste Zeichen muss ein Buchstabe sein.
- An den restlichen Positionen sind folgende Zeichen zugelassen: Buchstaben, Ziffern sowie die Symbole @, #, _ und $. Von der zweiten bis zur vorletzten Position ist außerdem der Punkt erlaubt.
- Aus den eben genannten Regeln ergibt sich insbesondere, dass Leerzeichen in Variablennamen verboten sind.
- Die von früheren SPSS-Versionen verschmähten Umlaute in Variablennamen werden neuerdings akzeptiert. Allerdings sind Probleme zu erwarten, wenn eine SPSS-Datendatei zu einem Rechner mit einem anderen Betriebssystem transferiert wird. Der unter MS-Windows vereinbarte Variablenname „größe“ wird z.B. auf dem Macintosh als „gr÷fle“ verstanden. Daher empfehle ich weiterhin, Umlaute und „ß“ in Variablennamen zu vermeiden.
- Die folgenden Schlüsselwörter der SPSS-Kommandosprache dürfen nicht als Variablennamen verwendet werden: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.
- Groß-/Kleinschreibung ist irrelevant.

Beim Versuch, einen irregulären Variablennamen zu vereinbaren, erhalten Sie im Dateneditor eine meist informative Fehlermeldung, z.B.:



Tipps zur Benennung:

- Bilden Sie möglichst *informative* Namen, also z.B. FNR, GESCHL und GEBJ für „Fallnummer“, „Geschlecht“ und „Geburtsjahr“ an Stelle unpraktischer Bezeichnungen wie VAR1, VAR2, VAR3.
- Die eben genannte Regel muss in einem speziellen Fall relativiert werden: Bei Serien verwandter Variablen (z.B. die 12 LOT-Fragen in Teil 3 unseres Erhebungsinventars) ist es in der Regel schwer, entsprechend viele individuelle Variablennamen zu bilden. Hier ist meist eine Indexschreibweise günstiger, bei der an einen informativen Namensstamm eine fortlaufende Nummer angehängt wird, z.B. LOT01, LOT02, ...

1.4.3.5 Kodierplan

Die Festlegungen zur Strukturierung und Kodierung der Projektdaten sollten in einem **Kodierplan** festgehalten werden. Er hat zwei Funktionen:

- Während der Erfassung regelt er, wie die Daten eines Falles ins Dateneditorfenster einzutragen bzw. mit einem anderen Programm zu erfassen sind.
- Später kann der Kodierplan als kompakte Beschreibung der entstandenen Datendatei verwendet werden.

Bei unserer KFA-Studie kann für die geplante Erfassung mit dem SPSS-Dateneditor z.B. der folgende Kodierplan verwendet werden:

Merkmal	SPSS-Var.-name	Kodierung	Bemerkungen
Fallnummer	FNR	MD-Indikator: entfällt	
Geschlecht	GESCHL	1 = Frau 2 = Mann MD-Indikator: SYSMIS	
Geburtsjahr	GEBJ	**zweistellige** Eingabe! MD-Indikator: SYSMIS	
Fachbereich	FB	1,..,7 MD-Indikator: SYSMIS	
Körpergröße	GROESSE	Eingabe in **cm**! MD-Indikator: SYSMIS	
Körpergewicht	GEWICHT	Eingabe in kg MD-Indikator: SYSMIS	
Ärger als Herr Meier (ohne KFA)	AERGO	0 = 0 1 = 10 . . . . . . 10 = 100 MD-Indikator: SYSMIS	
Ärger als Herr Schulze (mit KFA)	AERGM	0 = 0 1 = 10 . . . . . . 10 = 100 MD-Indikator: SYSMIS	
LOT-Fragen	LOT01 bis LOT12	1 = -- 2 = - 3 = o 4 = + 5 = ++ MD-Indikator: SYSMIS	
Kursmotive	MOTIV1 bis MOTIV5, KEINE	0 = nicht angekreuzt 1 = angekreuzt MD-Indikator: siehe Bem.	SYSMIS wird **nicht** vergeben. Die MD-Behandlung erfolgt später.
Spezielle Methoden gewünscht?	SMG	0 = nein 1 = ja MD-Indikator: SYSMIS	
Gewünschte statistische Methoden	METH1 bis METH5	1 = Meth.-Kat. 1 gew. . . . . . . MD-Indikator: Bei weniger als fünf Nennungen bleiben entsprechend viele Variablen **leer** (SYSMIS).	Das Kategorienschema wird während der Erfassung nach Bedarf entwickelt und dokumentiert. Eine differenzierte MD-Behandlung erfolgt später.

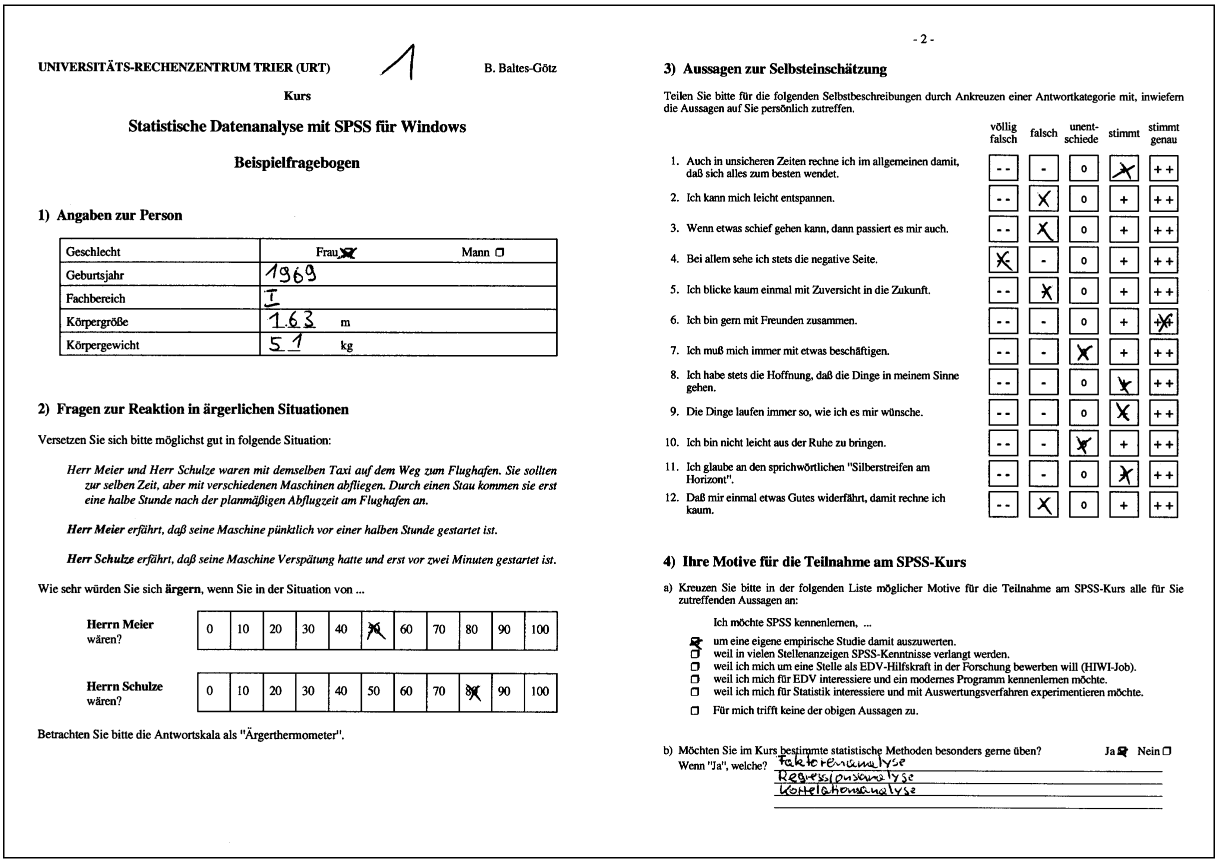
Dieser Kodierplan ist bei der Datenerfassung erfreulich einfach zu handhaben und leistet damit einen wichtigen Beitrag zur Integrität der auszuwertenden Daten.

1.5 Durchführung der Studie (inklusive Datenerhebung)

Bei den obigen Überlegungen zur Strukturierung und Kodierung der Daten hat sich ergeben, dass der in Abschnitt 1.3 wiedergegebene Fragebogen ohne Korrekturen eingesetzt werden kann. Damit steht der Durchführung unserer Befragung nichts mehr im Wege.

Im realen Kursverlauf übernehmen die Teilnehmer die Rolle der Probanden und produzieren so ihre eigenen, von zufälligen Stichprobeneffekten gefärbten Daten. Die Leser(innen) im Selbststudium werden wohl aus praktischen Gründen in der Regel auf die Durchführung einer eigenen KFA-Erhebung verzichten. Im weiteren Verlauf des Manuskriptes werden die in einem früheren Kurs erhobenen Daten analysiert. Die zugehörigen Dateien können über das Internet bezogen werden (siehe Vorwort).

Im folgenden ist der ausgefüllte Fragebogen der Untersuchungsteilnehmerin wiedergegeben, die bei der zufälligen Vergabe einer Fallidentifikation (vgl. Abschnitt 1.4.2.1) die Nummer 1 erhielt. Diese Nummer wurde nachträglich von der Untersuchungsleitung auf den Fragebogen geschrieben:



UNIVERSITÄTS-RECHENZENTRUM TRIER (URT) 1 B. Baltes-Götz

Kurs

Statistische Datenanalyse mit SPSS für Windows

Beispielfragebogen

1) Angaben zur Person

Geschlecht	Frau X Mann ☐
Geburtsjahr	1969
Fachbereich	I
Körpergröße	1.63 m
Körpergewicht	51 kg

2) Fragen zur Reaktion in ärgerlichen Situationen

Versetzen Sie sich bitte möglichst gut in folgende Situation:

Herr Meier und Herr Schulze waren mit demselben Taxi auf dem Weg zum Flughafen. Sie sollten zur selben Zeit, aber mit verschiedenen Maschinen abfliegen. Durch einen Stau kommen sie erst eine halbe Stunde nach der planmäßigen Abflugzeit am Flughafen an.

***Herr Meier** erfährt, daß seine Maschine pünktlich vor einer halben Stunde gestartet ist.*

***Herr Schulze** erfährt, daß seine Maschine Verspätung hatte und erst vor zwei Minuten gestartet ist.*

Wie sehr würden Sie sich **ärgern**, wenn Sie in der Situation von ...

Herrn Meier wären?	0	10	20	30	40	50 (X)	60	70	80	90	100
Herrn Schulze wären?	0	10	20	30	40	50	60	70	80 (X)	90	100

Betrachten Sie bitte die Antwortskala als "Ärgerthermometer".

- 2 -

3) Aussagen zur Selbsteinschätzung

Teilen Sie bitte für die folgenden Selbstbeschreibungen durch Ankreuzen einer Antwortkategorie mit, inwiefern die Aussagen auf Sie persönlich zutreffen.

	völlig falsch	falsch	unentschiede	stimmt	stimmt genau
1. Auch in unsicheren Zeiten rechne ich im allgemeinen damit, daß sich alles zum besten wendet.	- -	-	0	X	+ +
2. Ich kann mich leicht entspannen.	- -	X	0	+	+ +
3. Wenn etwas schief gehen kann, dann passiert es mir auch.	- -	X	0	+	+ +
4. Bei allem sehe ich stets die negative Seite.	X	-	0	+	+ +
5. Ich blicke kaum einmal mit Zuversicht in die Zukunft.	- -	X	0	+	+ +
6. Ich bin gern mit Freunden zusammen.	- -	-	0	+	X
7. Ich muß mich immer mit etwas beschäftigen.	- -	-	X	+	+ +
8. Ich habe stets die Hoffnung, daß die Dinge in meinem Sinne gehen.	- -	-	0	X	+ +
9. Die Dinge laufen immer so, wie ich es mir wünsche.	- -	-	0	X	+ +
10. Ich bin nicht leicht aus der Ruhe zu bringen.	- -	-	X	+	+ +
11. Ich glaube an den sprichwörtlichen "Silberstreifen am Horizont".	- -	-	0	X	+ +
12. Daß mir einmal etwas Gutes widerfährt, damit rechne ich kaum.	- -	X	0	+	+ +

4) Ihre Motive für die Teilnahme am SPSS-Kurs

a) Kreuzen Sie bitte in der folgenden Liste möglicher Motive für die Teilnahme am SPSS-Kurs alle für Sie zutreffenden Aussagen an:

Ich möchte SPSS kennenlernen, ...

- X um eine eigene empirische Studie damit auszuwerten.
- ☐ weil in vielen Stellenanzeigen SPSS-Kenntnisse verlangt werden.
- ☐ weil ich mich um eine Stelle als EDV-Hilfskraft in der Forschung bewerben will (HIWI-Job).
- ☐ weil ich mich für EDV interessiere und ein modernes Programm kennenlernen möchte.
- ☐ weil ich mich für Statistik interessiere und mit Auswertungsverfahren experimentieren möchte.
- ☐ Für mich trifft keine der obigen Aussagen zu.

b) Möchten Sie im Kurs bestimmte statistische Methoden besonders gerne üben? Ja X Nein ☐

Wenn "Ja", welche? Faktorenanalyse
Regressionsanalyse
Korrelationsanalyse

Vorsichtshalber möchte ich noch erwähnen, dass der Fragebogen hier aus Platzgründen stark verkleinert wiedergegeben wurde und in der Praxis keinesfalls in diesem Format vorgelegt werden darf.

2 Einstieg in SPSS für Windows

In den bisher dargestellten Projektphasen von der theoretischen Ausarbeitung bis zur Erstellung des Kodierplans spielte SPSS noch keine wesentliche Rolle. Die im KFA-Projekt nun anstehende Datenerfassung wollen wir jedoch mit diesem Programm bewerkstelligen, so dass an dieser Stelle einige einführende Bemerkungen zu SPSS und zu seiner Installation an der Universität Trier angemessen sind. In Abschnitt 2.1 lernen Sie, wie SPSS für Windows an der Universität Trier benutzt werden kann, und in den Abschnitten 2.2 bis 2.4 werden elementare Merkmale des Programms dargestellt.

2.1 SPSS für Windows an der Universität Trier

An der Universität Trier steht das Basis-System von SPSS für Windows mit den folgenden Erweiterungs-Modulen bzw. Zusatzprodukten zur Verfügung:

Erweiterungs-Module (in das Hauptprogramm integriert)	Zusatzprodukte (separat aufrufbar)
Regression Models Advanced Models Tables Trends Categories Conjoint Exact Tests	Amos Data Entry Answer Tree Neural Connection

Zu den Erweiterungs-Modulen finden Sie im Anhang jeweils eine kurze Beschreibung des Funktionsumfangs sowie Literaturangaben. Über die Zusatzprodukte informieren URT-Kursmanuskripte, die auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend folgendermaßen zu erreichen sind:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen

Die aufgeführten SPSS-Produkte können auf folgende Weise genutzt werden:

a) Pool-PCs unter MS-Windows

Auf den Pool-PCs unter dem Betriebssystem MS-Windows finden Sie über

Start > Programme

die Programmgruppe **SPSS vom NT-Server des URT** mit Unterverzeichnissen zu allen installierten SPSS-Produkten.

b) Erwerb einer Mietlizenz

Beschäftigte und Studierende der Universität Trier können über das URT eine befristete SPSS-Mietlizenz zur Verwendung im Rahmen ihrer dienstlichen Tätigkeit bzw. ihrer Ausbildung erwerben, wobei auch eine Installation im Privathaushalt erlaubt ist. Nähere Informationen erhalten Sie in der URT-Benutzerberatung.

2.2 Programmstart und Benutzeroberfläche

2.2.1 SPSS starten

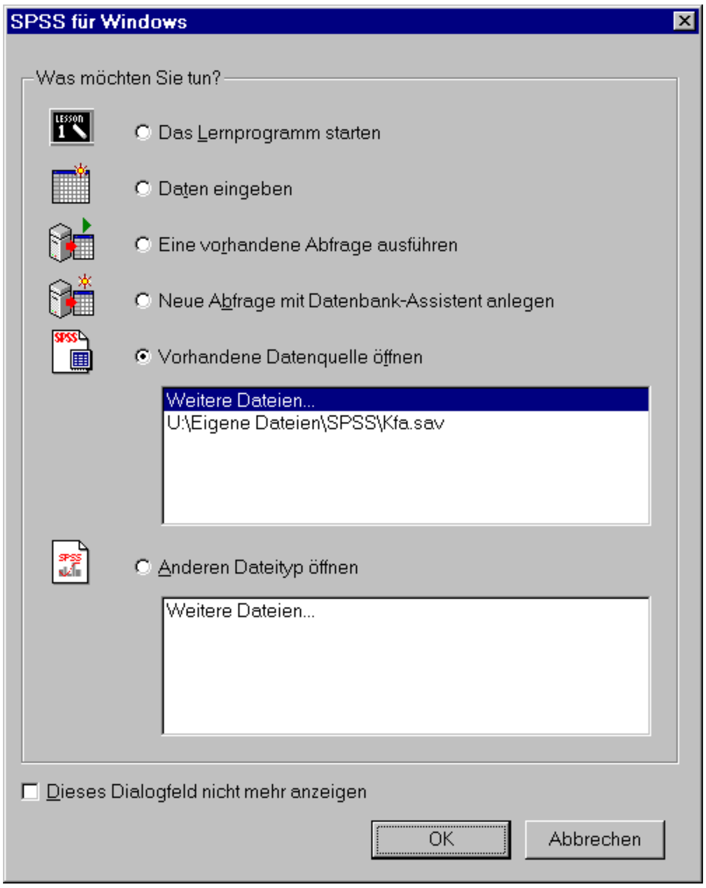
Nach erfolgreicher Anmeldung bei einem Pool-PC unter MS-Windows erreichen Sie SPSS 10 für Windows über das zugehörige Desktop-Symbol oder über das Startmenü:

Start > Programme > SPSS vom NT-Server des URT > SPSS 10.0 für Windows.

Auf einem PC mit lokaler SPSS-Installation können Sie das Programm in der Regel folgendermaßen starten:

Start > Programme > SPSS for Windows > SPSS 10.0 für Windows

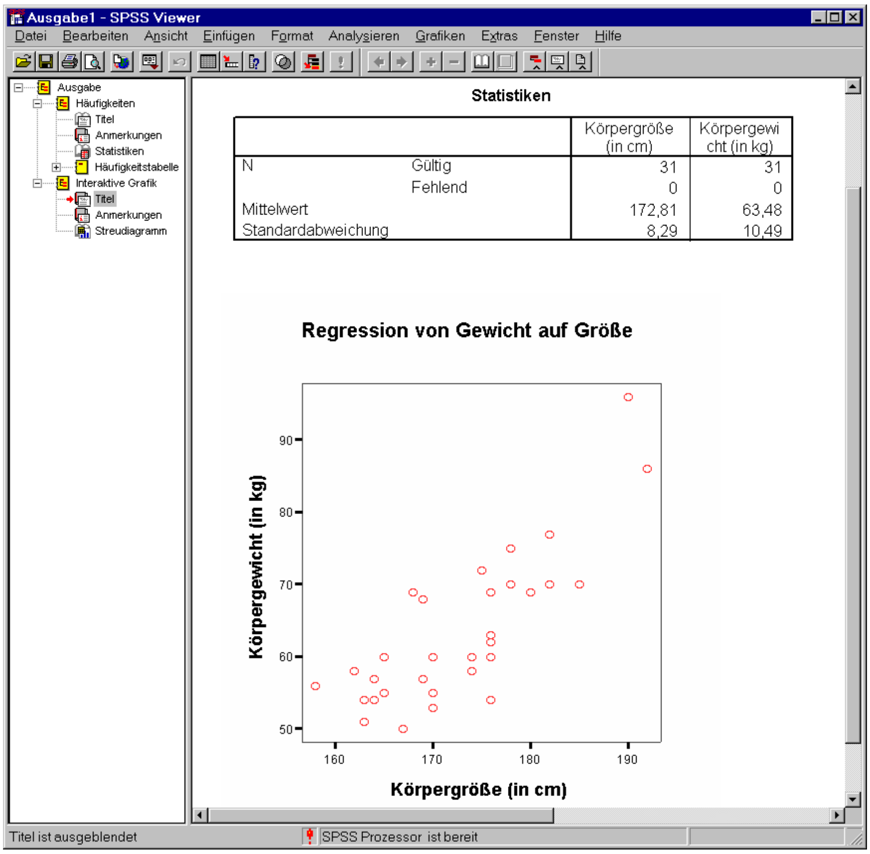
Nach dem Start erscheint der folgende Assistent:



Da wir seine Dienste im Kursverlauf nicht benötigen, können Sie seinen Auftritt über das Kontrollkästchen **Dieses Dialogfeld nicht mehr anzeigen** unterbinden.

2.2.2 Die wichtigsten SPSS-Fenster

Das Dateneditorfenster mit der (Fälle × Variablen) - Datenmatrix haben wir schon in Abschnitt 1.4.1 kennen gelernt. Nach der Datenerfassung können wir mit Hilfe seiner Menüzeile statistische und grafische Datenanalysen anfordern, die dann im **Ausgabefenster**, auch *SPSS Viewer* genannt, erscheinen, z.B.:



Die SPSS-Fenster enthalten in der Kopfzone eine Menüzeile und (verschiebbare) Symbolleisten, im Fußbereich eine Statuszeile mit Informationen über wichtige Programmzustände.

2.2.3 Was man mit SPSS so alles machen kann

Wir sind im Moment dabei, einen ersten Eindruck von dem „Arbeitsplatz“ SPSS für Windows zu gewinnen. Einen guten Überblick vermitteln die Optionen in der Menüzeile des Dateneditorfensters:

- **Datei**
  Hier finden Sie u.a. Befehle zum Öffnen bzw. Sichern von Datendateien sowie zum Beenden von SPSS.
- **Bearbeiten**
  Über das **Bearbeiten**-Menü erreichen Sie Editorbefehle zum Ausschneiden, Kopieren, Einfügen, Löschen und Suchen von Daten sowie die **Optionen**-Dialogbox zum Anpassen von SPSS-Eigenschaften. Außerdem können Sie hier manche Modifikationen des Datenfensters rückgängig machen.
- **Ansicht**
  Hier können Sie u.a. die Statuszeile sowie die Symbolleiste aus- bzw. einschalten sowie die Schriftart der angezeigten Daten festlegen.
- **Daten**
  Das **Daten**-Menü enthält u.a. Befehle zum Definieren von Variablen im Datenfenster, zum Zusammenfügen von SPSS-Dateien (z.B. mit Daten aus verschiedenen Stichproben) sowie zum Sortieren und Gewichten der Fälle.
- **Transformieren**
  Hier finden Sie z.B. die Befehle zum Rekodieren von Variablen oder zum Berechnen neuer Variablen aus bereits vorhandenen.

- **Analysieren**
  Dieser Menüpunkt erschließt die statistischen Auswertungsmethoden, mit denen wir letztlich unsere Forschungsfragen klären wollen.
- **Grafik**
  An dieser Stelle bietet SPSS vielfältige Möglichkeiten zur grafischen Präsentation von Daten an.
- **Extras**
  Hier finden sich diverse Funktionen (z. B. zur Anzeige von Informationen über die Variablen) sowie ein Menüeditor zum Anpassen der SPSS-Menüs.
- **Fenster**
  Über dieses Menü sind die aktuell offenen SPSS-Fenster erreichbar.
- **Hilfe**
  Hiermit starten Sie die Online-Hilfe, die Informationen über das gesamte SPSS-System bereithält und außerdem ein Lernprogramm sowie einen Statistik-Assistenten bietet.

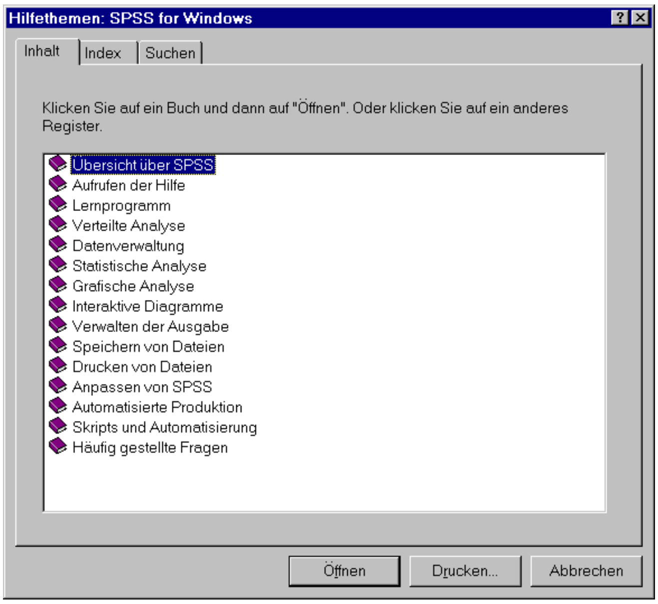
Bei leerem Datenfenster sind die meisten Menüoptionen nicht verfügbar. Die anderen SPSS-Fenster bieten angepasste Menüzeilen.

2.3 Hilfen an allen Ecken und Enden

Bei der Arbeit mit SPSS für Windows können Sie stets auf ein mächtiges Hilfesystem zurückgreifen, dessen wichtigste Möglichkeiten nun vorgestellt werden.

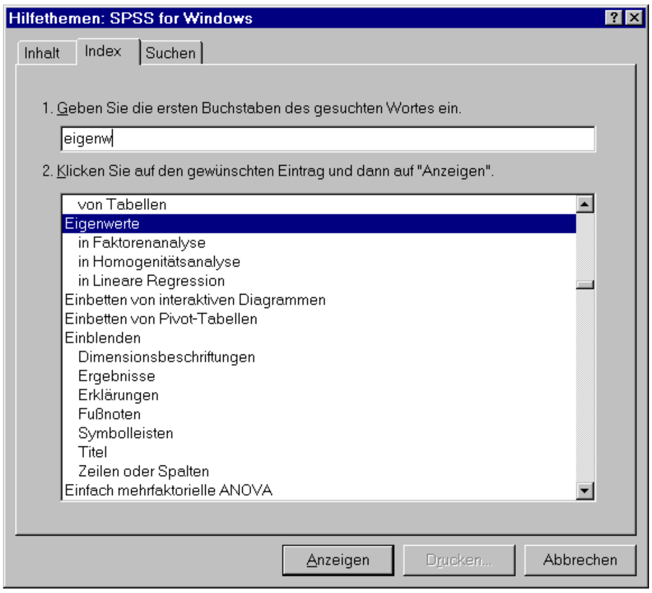
2.3.1 Das Online-Hypertext-Handbuch

Mit **Hilfe > Themen** erreichen Sie das Fenster **Hilfethemen**, das auf der Registerkarte **Inhalt** in systematischer Form über SPSS informiert:



2.3.2 Gezielte Suche nach SPSS-Begriffen

Die Registerblätter **Index** und **Suchen** im Fenster **Hilfethemen** eignen sich für die Suche nach Informationen zu bestimmten Begriffen, z.B.:



2.3.3 Das SPSS-Lernprogramm

Neben dem Fenster **Hilfethemen** mit seinen umfassenden Beschreibungen und seinem vollständigem Index gibt es ein weiteres Informationsangebot, das eher didaktisch orientiert und daher auf das Wichtigste beschränkt ist: das interaktive SPSS-Lernprogramm. Es wird mit **Hilfe > Lernprogramm** gestartet und sollte wie ein klassisches Einführungsbuch mehr oder weniger linear durchgearbeitet werden. In den einzelnen Kapiteln werden konkrete Arbeitsabläufe geübt, z.B.:



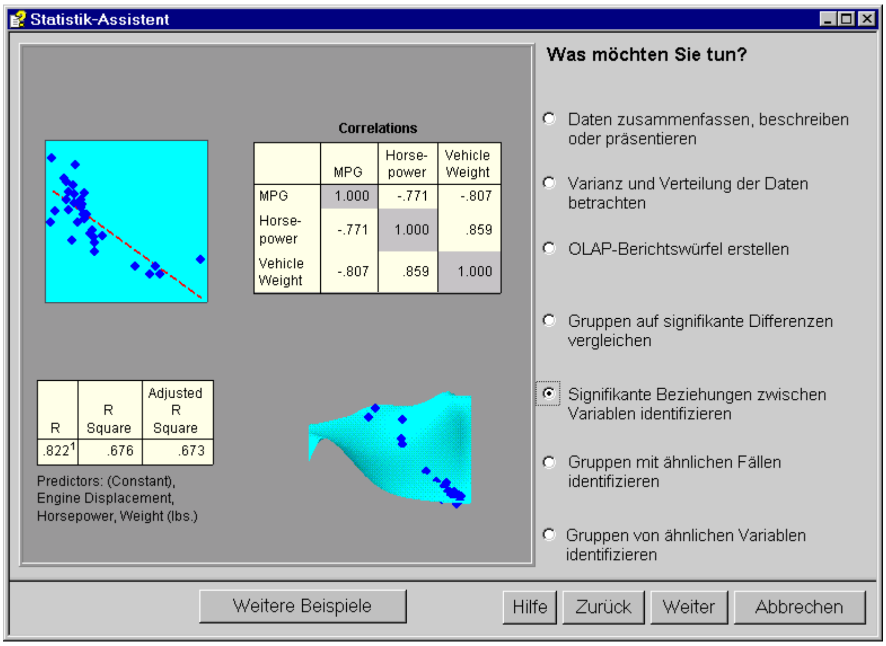
Sie können das Lernprogramm als eigenständige Windows-Anwendung parallel zu SPSS ausführen und damit die Lektionen sofort nachvollziehen, indem Sie zwischen SPSS und dem Lernprogramm hin und her wechseln (z.B. mit der Tastenkombination ALT ⇥).

2.3.4 Kontextsensitive Hilfe zu den Dialogboxen

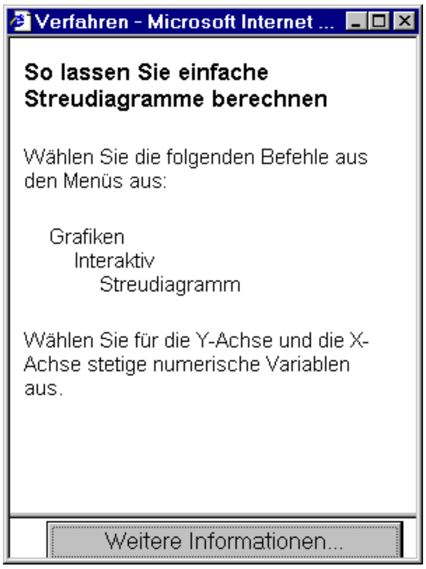
In jeder Dialogbox können Sie mit der Standardschaltfläche **Hilfe** Informationen zu all ihren Optionen anfordern.

2.3.5 Der Statistik-Assistent

Der über **Hilfe > Statistics Coach** verfügbare Assistent versucht, den statistisch weniger gut ausgebildeten Anwender durch eine Sequenz von Fragen zur richtigen Statistik- bzw. Grafikdialogbox zu führen, z.B.:



In einem einfachen Test mit dem oben abgebildeten Einstieg und der anschließenden Vereinbarung, dass zwei stetige Variablen grafisch analysiert werden sollen, hat der Assistent am Ende der Exploration tatsächlich das richtige Verfahren vorgeschlagen:



2.3.6 Syntax-Handbücher

Die Leistungen des SPSS-Systems lassen sich auch über eine mächtige Kommandosprache abrufen (siehe unten). Falls die Online-Syntaxhandbücher im PDF-Format installiert worden sind, können sie über den Menübefehl **Hilfe > Syntax Guide** aufgerufen werden.

2.3.7 SPSS im Internet

SPSS ist im Internet vielfach präsent, besonders zu erwähnen sind:

- **Die WWW-Homepage der SPSS Inc.:** http://www.spss.com
- **Die Usenet-Diskussionsgruppe comp.soft-sys.stat.spss**
  Hier werden technische und statistische Themen lebhaft diskutiert, wobei SPSS-Mitarbeiter zu wichtigen Fragen kompetent Stellung nehmen.

2.3.8 Benutzerberatung

Bei Problemen mit der Anwendung von SPSS können Sie sich an die Benutzerberatung des URT wenden.

Ort:	im Gebäude E (Räume 002a - 006).
Zeiten:	Montag bis Freitag: 10.30-11.30 Uhr, Montag bis Donnerstag: 14-16 Uhr

2.4 SPSS für Windows beenden

Die Beendigung einer SPSS-Sitzung wird mit

Datei > Beenden

eingeleitet. Falls Sie während der Sitzung Dokumente erstellt bzw. verändert und noch nicht gesichert haben (z.B. im Daten- oder im Ausgabefenster), werden Sie von SPSS an das Speichern erinnert.

3 Datenerfassung und der SPSS-Dateneditor

Bei einer typischen Studie liegen nach Abschluss der Datenerhebung für jeden Fall schriftliche Untersuchungsdokumente vor, die nun erfasst, d.h. in eine Computerdatei übertragen, werden müssen.

Allerdings müssen nicht nur Merkmalsausprägungen erfasst, sondern auch Variablenattribute deklariert werden (z.B. Namen, Indikatoren für fehlende Werte). Es hängt von der Erfassungsmethode ab, zu welchem Zeitpunkt und gegenüber welcher EDV-Komponente die Variablen deklariert werden.

Bevor in Abschnitt 3.2 die konkrete Erfassung der KFA-Daten mit dem SPSS-Dateneditor beschrieben wird, sollen in Abschnitt 3.1 einige alternative Erfassungsmethoden vorgestellt werden.

3.1 Methoden zur Datenerfassung

3.1.1 Automatisierte Verfahren

Zunächst geht es um zwei automatische Verfahren, die sich zunehmender Beliebtheit erfreuen.

3.1.1.1 Online-Datenerhebung

Wenn die nötigen technischen und organisatorischen Voraussetzungen gegeben sind, sollte eine **Online-Datenerhebung** eingesetzt werden. Hiermit sind Verfahren gemeint, bei denen die Untersuchungsteilnehmer(innen) ihre Daten (aktiv oder passiv) direkt in eine EDV-Anlage einspeisen (z.B. computergestützte Diagnostik, automatische Aufzeichnung physiologischer Daten). Nach Abschluss der Datenerhebung kann sofort die Datenauswertung beginnen, wobei jedoch eine Prüfung der Daten nach unterschiedlichen Gesichtspunkten in der Regel nötig ist (z.B. wegen möglicher Defekte in der Aufzeichnungsapparatur). Die Daten*erfassung* als eigenständige Arbeitsphase entfällt bei den Online-Verfahren. Mit dem Aufkommen einer verbreiteten Netzwerkinfrastruktur (z.B. Internet), wird es möglich sein, Online-Befragungen mit repräsentativen Stichproben durchzuführen.

3.1.1.2 Automatisches Einscannen von schriftlichen Untersuchungsdokumenten

Auch nach einer schriftlichen Befragung im konventionellen Stil lässt sich das manuelle Erfassen der Daten vermeiden. Diese lästige und fehleranfällige Arbeit kann man einer EDV-Anlage zum automatischen Einscannen und Interpretieren der schriftlichen Untersuchungsdokumente übertragen. Allerdings muss die EDV-Anlage erst mit einigem Aufwand in ihre Arbeit eingewiesen werden, so dass bei kleineren Projekten kaum ein Rationalisierungsgewinn zu erzielen ist. Eine weitere Voraussetzung dieses Verfahrens ist die Beachtung einiger Regeln beim Entwurf der Untersuchungsmaterialien. Insgesamt gesehen ist das Einscannen von Fragebögen sicher für viele Forschungsprojekte eine attraktive und rentable Erfassungsmethode. An der Universität Trier stehen für diesen Zweck im Grafikraum des Rechenzentrums (E-020) folgende Windows-Programme mit der erforderlichen Hardware (Scanner mit automatischem Einzelblatteinzug) zur Verfügung:

- **PaperKeyboard**
  PaperKeyboard 97 OCR Server 3.0 kann Markierungen (OMR) und gedruckte Zeichen (OCR) lesen. Der zu erfassende Fragebogen wird mit einem beliebigen Textverarbeitungsprogramm erstellt, wobei allerdings Designrestriktionen zu berücksichtigen sind. Anschließend folgt die Formulardeklaration und schließlich die Datenerfassung. Damit können bereits vorhandene Fragebögen oft nach einigen Modifikation weiter verwendet werden.

Auch lassen sich bei der aktuellen PaperKeyboard-Version die technischen Anforderungen an die Fragebögen in der Regel unauffällig in das Design integrieren.
Beim Datenexport wird SPSS bedauerlicherweise nicht direkt unterstützt, so dass man einen Umweg über ein Fremdformat wählen muss (z.B. Textdatei mit tabulatorbegrenzten Werten).

- **Teleform**
  Teleform 6.1 kann neben Markierungen (OMR) und gedruckten Zeichen (OCR) auch Handschrift lesen (ICR), wobei die Handschrifterkennung m.E. mit Vorsicht eingesetzt werden sollte. Im Unterschied zu PaperKeyboard enthält Teleform einen Formulargenerator, so dass Fragebogendesign und -deklaration in einem Arbeitsschritt erfolgen. Allerdings ist der eingebaute Formulargenerator nicht immer so flexibel und bequem zu bedienen wie das gewohnte Textverarbeitungsprogramm, was das Fragebogendesign in der Regel erschwert.
  Beim Datenexport kann Teleform direkt in eine SPSS-Datendatei schreiben und dabei seine Feldbezeichnungen als SPSS-Variablennamen übergeben.

Ausführliche Beschreibungen zu PaperKeyboard bzw. Teleform finden Sie im Intranet der Universität Trier von der Startseite (www.uni-trier.de) ausgehend über:

Weitere Service-Angebote > EDV-Dokumentationen > Elektronische Publikationen > Datenerfassung

3.1.2 Manuelle Verfahren

Im restlichen Teil dieses Abschnitts werden wir uns mit den noch immer dominierenden manuellen Erfassungsmethoden beschäftigen, bei denen die Daten gemäß Kodierplan „in den Computer eingetippt“ werden.
Zunächst einige Empfehlungen, die für alle manuellen Erfassungsmethoden gelten:

- Schon beim Entwurf des Kodierplans (siehe oben) ist darauf zu achten, dass dem Erfasser keine unnötigen und fehleranfälligen Arbeiten zugemutet werden, z.B.:
  - Häufig sind nicht die erfassten Rohvariablen selbst, sondern darauf aufbauende Variablen (z.B. Mittelwerte) von Interesse. SPSS verfügt über leistungsfähige Befehle zur Berechnung neuer Variablen aus bereits vorhandenen, so dass man derartige Routinearbeiten keinesfalls während der Datenerfassung „zu Fuß“ ausführen sollte.
  - Wurden einige Fragen aus meßtechnischen Gründen umgepolt (negativ formuliert), so sollten Sie diese Umpolung *nicht* während der Erfassung rückgängig gemacht. Dies geht sehr viel bequemer und ohne Fehlerrisiko mit den Transformationsmöglichkeiten von SPSS (siehe unten).
- Übertragen Sie Daten von Fragebögen oder ähnlichen Untersuchungsmaterialien *direkt* in den Rechner. Das gelegentlich empfohlene Verfahren, die Daten zunächst von den Untersuchungsdokumenten auf sogenannte Kodierbögen zu übertragen, um sie dann von dort endgültig zu erfassen, erhöht den Aufwand und die Fehlerwahrscheinlichkeit.

Von den möglichen manuellen Erfassungsmethoden sollen drei in diesem Manuskript vorgestellt werden:

- Erstellung einer Text-Datendatei mit einem beliebigen Texteditor
  Die Erfassung in eine Text-Datendatei hat nur einen einzigen Vorteil: Man kann sie mit fast jedem beliebigen Texteditor durchführen, z.B. auch mit dem vertrauten Textverarbeitungsprogramm. Ihr wesentlicher Nachteil ist die hohe Fehleranfälligkeit (vgl. Abschnitt 4). Diese veraltete Erfassungsmethode wird hier nur beschrieben, um Sie davon

abzuhalten. Allerdings gibt es noch einen zweiten Grund, das Innenleben von Text-Datendateien zu beschreiben: Es sind noch sehr viele Exemplare im Umlauf, die Sie eventuell auswerten müssen. Daher kommen wir nicht umhin, später das Einlesen von Text-Datendateien zu behandeln.

- Erfassung mit dem SPSS-Dateneditor
  Der SPSS-Dateneditor ist ein integraler Bestandteil des SPSS-Systems, so dass wir uns mit seiner Bedienung auf jeden Fall vertraut machen müssen. Er ist nicht perfekt optimiert für die Erfassung größerer Datenmengen, kann aber in kleinen bis mittleren Projekten verwendet werden.
- Einsatz eines speziellen Datenerfassungsprogramms
  Ein spezielles Datenerfassungsprogramm bietet Vorteile gegenüber dem SPSS-Dateneditor, erfordert aber auch zusätzlichen Einarbeitungsaufwand.

Aufgrund des relativ geringen Datenaufkommens in unserem KFA-Projekt ist der SPSS-Dateneditor die optimale Erfassungsmethode. Weil in Abschnitt 3.2 die Erfassung der KFA-Daten mit dem SPSS-Dateneditor ausführlich beschrieben wird, müssen im aktuellen Abschnitt nur die beiden anderen manuellen Erfassungsmethoden vorgestellt werden.

3.1.2.1 Erstellung einer Text-Datendatei mit einem beliebigen Texteditor

Bei dieser veralteten, zeitaufwendigen und vor allem sehr fehleranfälligen Methode muss festgelegt werden, wie die Beobachtungswerte eines Falles in der Textdatei angeordnet werden sollen. Im wesentlichen stehen zwei Alternativen zur Auswahl: positionierte Daten und separierte Daten.

Positionierte Daten

In einer Datei mit fest positionierten bzw. formatierten Daten beginnt der Datensatz jedes Falles auf einer neuen Datenzeile. Ferner hat jede Variable einen festen Standort im Datensatz eines Falles (z.B. in Zeile 1, Spalten 12-13). Damit sind die Datensätze aller Fälle identisch aufgebaut. So sehen die festformatig per Texteditor erfassten KFA-Daten aus, die wir im Manuskript analysieren werden[1]:

[1] Da unser Kodierplan für die Erfassung per SPSS-Dateneditor konzipiert ist, enthält er keine Zeilen-Spalten-Positionen für die Variablen. Diese wurden eigens für die Erstellung der Daten-Textdatei festgelegt. Dies geschah im Rahmen des folgenden SPSS-Programms, welches die Textdatei (nachträglich) über das WRITE-Kommando aus der vorhandenen SPSS-Datendatei erstellt hat:

```
write outfile='kfar.txt'
 /fnr 1-2 '1' geschl 5 gebj 6-7 fb 8 groesse 9-11 gewicht 12-13
 /fnr 1-2 '2' aergo aergm 5-8 lot01 to lot12 10-21 motiv1 to keine 23-28
  smg 30 meth1 to meth5 31-40.
exe.
```

Unser Kodierplan sieht außerdem die systematische Verwendung des MD-Indikators SYSMIS vor. Dies ist jedoch bei Text-Datendateien nicht sinnvoll. Hier sollten benutzerdefinierte MD-Indikatoren verwendet werden.

```
 11 169116351
 12   5 8 422125344342 100000 1 1 2 3 0 0
 21 170115856
 22   5 8 431224342342 100000 1 1 2 0 0 0

        .
        .
        .

301 167117060
302   910 551115443131 100000 0 0 0 0 0 0
311 167116968
312   7 9 412544231132 100010 1 1 3 0 0 0
```

Separierte Daten

In einer Datei mit separierten Daten müssen die Variablenausprägungen jedes Falles in derselben Reihenfolge vorliegen, und je zwei Werte müssen durch ein Separatorzeichen voneinander getrennt werden. Beim Trennzeichen hat man die freie Auswahl, entscheidet sich aber meist zwischen folgenden Kandidaten:

- Tabulatorzeichen
- Komma
- Semikolon
- Leerzeichen

Beim Einlesen separierter Daten durch SPSS wird eine Serie aufeinander folgender Leerzeichen behandelt wie ein einzelnes Leerzeichen. Ansonsten schließen zwei aufeinaderfolgende Trennzeichen einen fehlenden Wert ein, den SPSS beim Einlesen als SYSMIS interpretiert.
Obwohl nicht zwingend vorgeschrieben, sollte man alle Daten eines Falles in eine einzige Zeile schreiben und für jeden Fall eine neue Zeile beginnen.
In der ersten Zeile einer Textdatei mit separierten Daten können die Variablennamen an SPSS übergeben werden, was im folgendem Beispiel mit Tabulator-separierten KFA-Daten demonstriert wird:

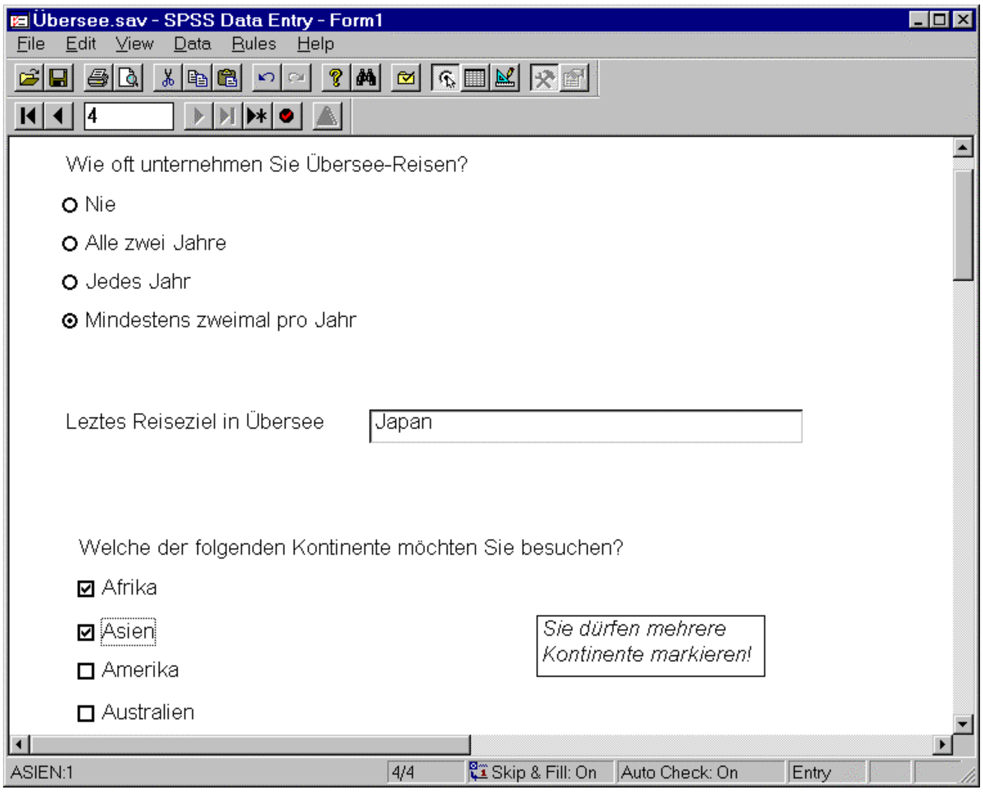
```
FNR   GESCHL      GEBJ  FB    GROESSE     GEWICHT     AERGO AERGM .  .  .
1     1     69    1     163   51    5     8     .  .  .
2     1     70    1     158   56    5     8     .  .  .
3     1     69    1     174   58    4     8     .  .  .
4     2     67    1     182   77    6     2     .  .  .
5     1     67    1     180   69    8     8     .  .  .
.     .     .     .     .     .     .     .
.     .     .     .     .     .     .     .
29    1     68    1     176   63    7     9     .  .  .
30    1     67    1     170   60    9     10    .  .  .
31    1     67    1     169   68    7     9     .  .  .
```

3.1.2.2 Einsatz eines speziellen Datenerfassungsprogramms

Wenn bei größeren Projekten eine manuelle Datenerfassung unumgänglich ist (vgl. Abschnitt 3.1.1), dann sollte in der Regel ein spezielles Datenbankprogramm verwendet werden. Man arbeitet hier bequem mit einer *Erfassungsmaske*, die einen einzelnen Fall in übersichtlicher Form auf dem Bildschirm präsentiert. Zudem werden die eingegebenen Daten in der Regel *sofort auf Plausibilität überprüft*: Falsche Eingaben werden mit entsprechendem Protest abgewiesen.
Ein Nachteil dieser Methode besteht darin, dass ein spezielles Programm erlernt werden muss. Nach der Lernphase ist jedoch im Vergleich zu den anderen manuellen Erfassungsmethoden mit dem geringsten Zeitbedarf und der kleinsten Fehlerquote zu rechnen.
Von den zahlreichen Datenbankprogrammen, die zur Datenerfassung verwendet werden können, sollen zwei exemplarisch beschrieben werden:

a) SPSS Data Entry

Mit **Data Entry** kann man eine analog zum Fragebogen aufgebaute Eingabemaske entwerfen, um dem Erfasser die Orientierung zu erleichtern, z.B.:



Es ist allerdings zu befürchten, dass derart locker verteilte Eingabebereiche die Erfassungsgeschwindigkeit beeinträchtigen.
Die folgenden Data Entry – Vorzüge sind jedoch unbestreitbar:

- **Filterfragen** („jump-and-fill-rules")
  In Abhängigkeit vom erfassten Wert einer Filtervariablen verzweigt Data Entry zu unterschiedlichen Folgevariablen und versorgt dabei übersprungene Variablen mit einem festgelegten MD-Indikator.
- **Plausibilitätsprüfungen**
  Man kann z.B. dafür sorgen, dass bei der Variablen GESCHL nur die Werte 0, 1 und SYSMIS eingegeben werden können.

Neben der Daten*erfassung* will Data Entry auch das Fragebogen*design* unterstützen. Man kann entweder *ein* Formular zur Verwendung bei der Datenerhebung (z.B. durch schriftliche Befragung) *und* bei der EDV-Erfassung entwerfen, oder für beide Anwendungsfälle angepasste Formulare verwenden. Dazu bietet Data Entry Beispielfragebögen bzw. Musterbibliotheken (z.B. mit demographischen Fragen) an.
Weitere Funktionen von Data Entry sind:

- Existierende SPSS-Datendateien auf Fehler prüfen
- Einen Fragebogen zu einer existierenden SPSS-Datendatei erstellen

Eine ausführlich Beschreibung zu Data Entry finden Sie auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend über:

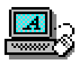
Weitere Service-Angebote > EDV-Dokumentationen > Elektronische Publikationen > Datenerfassung

b) INPUT II

Ein anderer Weg zum maßgeschneiderten Datenbankprogramm mit maskengesteuerter Dateneingabe, Plausibilitätskontrolle und Filterführung ist die Verwendung des Programmgenerators **INPUT II**, der an der Universität Trier für die speziellen Bedürfnisse wirtschafts- und sozialwissenschaftlicher Forschungsarbeit mit SPSS entwickelt wurde. Der (nur im Campusnetz den Universität Trier verfügbare) Generator erzeugt aus einer Datensatzbeschreibung ein spezialisiertes Erfassungsprogramm, das dann ohne Lizenzgebühren auf jedem beliebigen PC (z.B. auch zu Hause) eingesetzt werden kann. Das DOS-basierte Erfassungsprogramm ist zwar nicht mehr ganz auf der Höhe moderner Softwaretechnik, begnügt sich dafür aber auch mit einer minimalen Hardwareausstattung.
Eine INPUT II - Beschreibung finden Sie auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend über:

Weitere Service-Angebote > EDV-Dokumentationen > Elektronische Publikationen > Datenerfassung

3.2 *Erfassung mit dem SPSS-Dateneditor*



Für die nächsten Schritte im KFA-Projekt benötigen wir eine SPSS-Sitzung mit einem leeren Datenfenster. Dies ist z.B. der Fall, nachdem Sie SPSS gestartet und ggf. den aufdringlichen Startassistenten mit dem Ziel **Daten eingeben** verlassen haben. Nötigenfalls können Sie ein aktives und leeres Datenfenster mit dem folgenden Menübefehl anfordern:

Datei > Neu > Daten

Im realen SPSS-Kurs werden wir nun mit dem SPSS-Dateneditor unsere Variablen deklarieren und anschließend die Daten erfassen.
Wenn Sie dieses Manuskript im Selbststudium lesen, können und sollten Sie trotzdem die folgenden Arbeitsschritte zur Variablendefinition konkret nachvollziehen und die Daten des im Manuskript abgedruckten ersten Falles eintragen (siehe Seite 21). Alle Projektphasen nach der Datenerfassung können Sie durch Verwendung der SPSS-Datendatei **kfar.sav** mitmachen, deren Inhalt im weiteren Verlauf erklärt wird. Wie Sie diese Datei von einem Server des Rechenzentrums beziehen können, wird im Vorwort erklärt.

3.2.1 Dateneditor und Arbeitsdatei

Wir haben schon in Abschnitt 1.4.1 festgestellt, dass über das Dateneditorfenster[1] die rechteckige (Fälle ×Variablen) - Datenmatrix zugänglich ist. SPSS speichert die Daten während der Sitzung in einer temporären Datei, bezeichnet als **Arbeitsdatei** oder **Arbeitsdatendatei**, die nach Möglichkeit im Hauptspeicher des PCs gehalten wird. Die im Dateneditorfenster sicht- und modifizierbare Arbeitsdatei ist gewissermaßen die zentrale Datenbasis, die während einer SPSS-Sitzung aufgebaut, verändert und natürlich auch bei Statistik bzw. Grafikanforderungen verwendet wird. Sie enthält:

[1] Wie Sie sicher schon bemerkt haben, wird im Manuskript gelegentlich für „Dateneditorfenster“ die kürzere Bezeichnung „Datenfenster“ verwendet.

- Die **rechteckige (Fälle × Variablen)-Datenmatrix**
  Wir wollen statistische Analysen für die Variablen anfordern, d.h. für die Spalten der (Fälle × Variablen)-Datenmatrix in der Arbeitsdatei. Dazu ist jede Variable über ihren eindeutigen Variablennamen ansprechbar.
- Einen sogenannten **Deklarationsteil**
  Dort merkt sich SPSS verarbeitungsrelevante Merkmale der Variablen (z.B. MD-Indikatoren). Über die **Variablenansicht** des Datenfensters (siehe unten) können Sie die Merkmale der Variablen jederzeit einsehen und ändern.

Mit Hilfe des Dateneditors oder durch Transformationskommandos (siehe unten) können während einer Sitzung u.a. folgende Modifikationen der Arbeitsdatei vorgenommen werden:

- Erweiterung um neue Variablen
- Änderung von Variablenattributen (z.B. Namen, MD-Indikatoren)
- Löschen von Variablen
- Erweiterung um neue Fälle
- Änderung von Variablenausprägungen eines Falles
- Löschen von Fällen

Weil die Begriffe **Dateneditor** und **Arbeitsdatei** für den Umgang mit SPSS recht wichtig sind, sollen ihre wesentliche Eigenschaften noch einmal wiederholt werden:

- Die Arbeitsdatei enthält die Datenmatrix und den Deklarationsteil.
- Mit dem Dateneditor können wir die Arbeitsdatei ansehen und modifizieren, auf dem Registerblatt **Datenansicht** die Datenmatrix und auf dem Registerblatt **Variablenansicht** den Deklarationsteil.
- Die Arbeitsdatei ist temporär, muss also nach einer (planvollen) Änderung in eine permanente SPSS-Datendatei gesichert werden (siehe unten).

3.2.2 Variablen definieren

Wie eben erwähnt, verwaltet SPSS für jede Variable zahlreiche verarbeitungsrelevante Merkmale (z.B. MD-Indikatoren). Diese werden im Deklarationsteil der Arbeitsdatei gespeichert und können vom Benutzer bestimmt werden. Da SPSS für alle Attribute geeignete Voreinstellungen benutzt, setzt die Datenerfassung nicht unbedingt eine Variablendefinition voraus[1], doch wird das Erfassen z.B. dadurch erleichtert, dass die benutzerdefinierten Variablennamen über den Spalten der Datenmatrix stehen anstelle der voreingestellten Namen VAR00001, VAR00002 usw. Daher liegt es nahe, dem SPSS-System die in unserem Kodierplan beschriebenen Variablen vor dem Eintragen der Daten bekannt zu machen.

3.2.2.1 Das Datenfenster-Registerblatt Variablenansicht

Seit der SPSS-Version 10 besitzt das Datenfenster *zwei* Registerblätter bzw. Tabellen:

- das Registerblatt **Datenansicht** zur Anzeige und Modifikation der (Fälle × Variablen)-Datenmatrix
- das Registerblatt **Variablenansicht** zur Anzeige und Modifikation der Variablenattribute

[1] Da in SPSS der Variablentyp *numerisch* voreingestellt ist, müssten wir vor dem Erfassen von Daten anderen Typs auf jeden Fall eine Variablendefinition vornehmen. Allerdings sind solche Variablen in unserem Kodierplan nicht vorgesehen.

In den Zeilen der **Variablenansicht** wird jeweils eine Variable beschrieben, wozu in den Spalten insgesamt 10 Attribute zur Verfügung stehen. Für unsere erste Variable (FNR) eignen sich z.B. folgende Angaben:



Um eine neue Variable anzulegen, trägt man ihre Attribute in eine freie Zeile der Tabelle ein, wobei das Eintragen der einzelnen Attribute durch angepasste Dialogelemente unterstützt wird. Darüber hinaus können auch Variablen eingefügt, gelöscht oder verschoben werden (siehe unten).

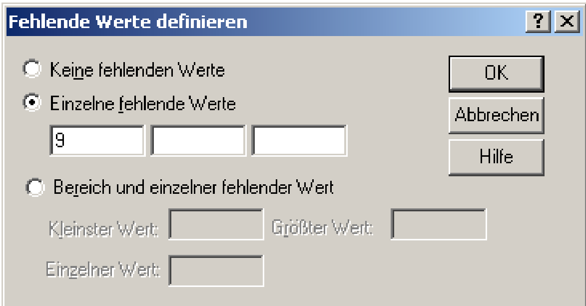
3.2.2.2 Die SPSS-Variablenattribute

Bevor wir die Variablen unsere KFA-Studie deklarieren, sollen vorab die SPSS-Variablenattribute erläutert werden:

- **Name**
  Die wesentlichen Regeln für SPSS-Variablennamen wurden schon im Zusammenhang mit dem Kodierplan genannt (siehe Seite 19).
- **Typ**
  Die wichtigsten SPSS-Variablentypen haben wir schon genannt: Numerisch, String und Datum (siehe Seite 13). In der Regel empfiehlt es sich, bei allen Variablen (auch bei den nominalskalierten) eine numerische Kodierung vorzunehmen (siehe Abschnitt 1.4.3), so dass der voreingestellte Variablentyp beibehalten werden kann.
- **Spaltenformat** und **Dezimalstellen**
  Bei einer **numerischen** Variablen können Sie festlegen, wie deren Werte im Datenfenster bzw. in der Ergebnisausgabe *angezeigt* werden sollen:
  Spaltenformat = Anzahl der anzuzeigenden Stellen
  Dezimalstellen = Anzahl der anzuzeigenden Dezimalstellen
  Diese Angaben betreffen *nicht* die Speichergenauigkeit im Datenfenster bzw. in der Arbeitsdatei, sondern nur die Anzeige.
  Bei einer **alphanumerischen** Variablen ist mit dem Spaltenformat keine „oberflächliche„ Anzeige-Eigenschaft gemeint, sondern die tatsächlich Maximalbreite der Variablen im Speicher. So werden z.B. bei einer nachträglichen Reduktion der Spaltenzahl tatsächlich entsprechend viele Zeichen (am rechten Rand) gelöscht.
- **Variablenlabel** und **Wertelabels**
  Hier können optional Variablenlabel zur näheren Erläuterung der (auf acht Zeichen beschränkten) Variablennamen sowie Wertelabels zur Erläuterung von Variablenausprägungen vereinbart werden. Diese Labels spielen bei Berechnungen keine Rolle, erscheinen aber in der Ergebnisausgabe und können deren Lesbarkeit verbessern.
  Sind Variablenlabel vorhanden, werden diese auch in Dialogboxen zur Beschreibung der Variablen verwendet. Diese Voreinstellung kann aber über
  Bearbeiten > Optionen > Allgemein > Variablenlisten = Namen anzeigen

abgeändert werden. Bei der in Dialogboxen üblichen Platzbeschränkung auf ca. 20 Stellen ist oft der abgeschnittene Anfang eines 50-stelligen Labels weniger informativ als der vollständige (achtstellige) Name.

- **Fehlende Werte**
  Wenn Sie bei einer Variablen *benutzerdefinierte* MD-Indikatoren verwendet haben, müssen Sie diese hier deklarieren, weil sie sonst wie gültige Werte verrechnet werden, z.B. bei einer Mittelwertsbildung.
  Da wir im KFA-Projekt laut Kodierplan ausschließlich System-Missing als MD-Indikator verwenden, müssen wir anschließend keine MD-Deklaration vornehmen (vgl. Abschnitt 1.4.3.5). Daher wird an dieser Stelle kurz die simple Prozedur zum Deklarieren von benutzerdefinierten MD-Indikatoren beschrieben:
    - Markieren Sie die Zelle zum Attribut **Fehlende Werte**.
    - Nach einem Mausklick auf den nun vorhandenen Erweiterungsschalter ... erscheint eine Dialogbox, in der u.a. bis zu 3 einzelne Werte als MD-Indikatoren vereinbart werden können, z.B.:



- **Spalten** und **Ausrichtung**
  Wie breit soll die Spalte einer Variablen im Datenfenster sein? Wie sollen die Werte ausgerichtet werden (linksbündig, zentriert, rechtsbündig)? Die Attribute in dieser Subdialogbox wirken sich nur auf die Darstellung einer Variablen im Datenfenster aus.

- **Messniveau**
  Über die technischen Variablenattribute hinaus kann seit der SPSS-Version 8 das Messniveau einer Variablen festgelegt werden, wobei diese Vereinbarung bei der weiteren Arbeit mit SPSS allerdings bisher nur in wenigen Situationen relevant ist:
    - Beim Öffnen von SPSS-Datendateien mit **AnswerTree**
    - Bei manchen Diagrammtypen (z.B. Streuungsdiagramm, siehe unten)

  In Zukunft werden wohl mehr SPSS-Prozeduren die Information über das Messniveau der Variablen ausnutzen. Weil außerdem die Reflexion über dieses methodologisch wichtige Merkmal der Variablen keinesfalls schaden kann, wollen wir uns in diesem Kurs der Pflicht unterziehen, bei allen Variablen das korrekte Messniveau anzugeben.

3.2.2.3 Variablendefinition durchführen

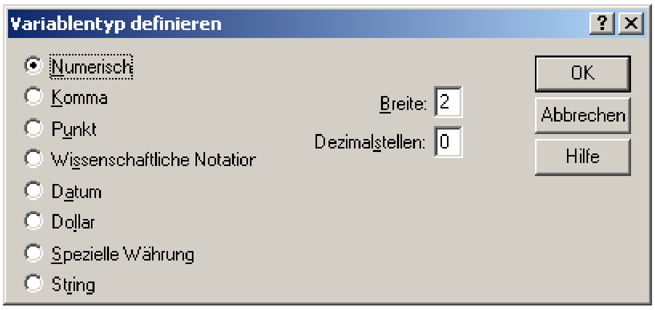
Aktivieren Sie nun die **Variablenansicht** des Datenfensters, und tragen Sie für die erste Variable (zur Fallidentifikation) den Namen FNR ein. Nach dem Markieren der zugehörigen Zelle können Sie sofort mit dem Eintippen des Namens beginnen. Die Groß/Kleinschreibung ist dabei irrelevant. Im Manuskript werden Variablennamen nur aus darstellungstechnischen Gründen groß geschrieben.
Sobald Sie die Zelle verlassen (z.B. durch Markieren einer anderen Zelle) wird eine neue Variable mit dem gewünschten Namen in die Arbeitsdatei aufgenommen, sofern gegen den Variablennamen keine Einwände bestehen. Die restlichen Attribute der neuen Variablen werden mit Standardwerten versorgt.

Nach dem Markieren der Zelle **Spaltenformat** kann die gewünschte Anzeigebreite durch Eingeben einer Zahl oder über eine vertikale Bildlaufleiste in Minimalausführung festgelegt werden:



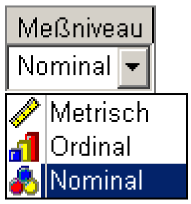
Analog wird auch die Anzahl der **Dezimalstellen** festgelegt.

Eine alternative Möglichkeit zum Einstellen der beiden Attribute findet sich in der (von uns eigentlich nicht benötigten) Dialogbox **Variablentyp**, die über einen Mausklick auf den Erweiterungsschalter ··· in der markierten **Typ**-Zelle erscheint:

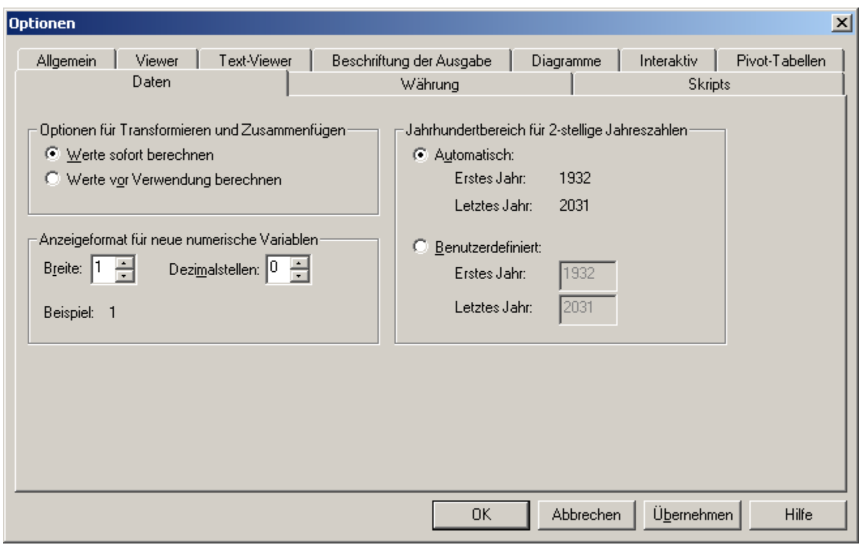


Im Datenfenster benötigt die Variable FNR nicht mehr als 3 **Spalten**, damit Werte und Namen vollständig sichtbar sind.

Klappen Sie schließlich in der markierten **Messnivau**-Zelle die versteckte Liste auf, um ein nominales Skalenniveau zu deklarieren:

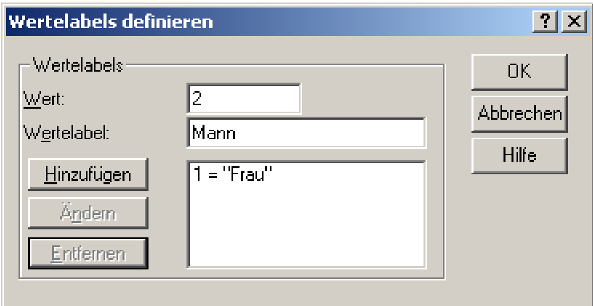


Tipp: Wenn in einem Projekt das voreingestellte Anzeigeformat für numerische Variablen (Breite = 8, Dezimalstellen = 2) häufig durch eine bestimmte Alternative ersetzt werden muss, kann zur Vereinfachung der Deklaration die Voreinstellung entsprechend geändert werden. Dazu öffnet man mit **Bearbeiten > Optionen** die Dialogbox **Optionen**, wechselt hier zum Registerblatt **Daten** und nimmt im Rahmen **Anzeigeformat für neue numerische Variablen** die gewünschten Einstellungen vor, z.B.:



MD-Indikatoren müssen wir im KFA-Projekt generell nicht vereinbaren, Variablen- und Wertelabels sind bei der augenblicklich bearbeiteten Fallnummernvariablen irrelevant. Daher können wir die Definition der Variablen FNR vorläufig beenden. Bei Bedarf sind Anpassungen jederzeit möglich.

Vereinbaren Sie nun in der zweiten Zeile der Variablenansicht für die Geschlechtsvariable den Namen GESCHL, eine einspaltige Anzeige ohne Dezimalstellen und nominales Messniveau. Tragen Sie das Variablenlabel „Geschlecht" in die zugehörige Zelle ein. Es wird im Ergebnisprotokoll ergänzend zum Variablennamen ausgegeben.
Bei der nominalskalierten Geschlechtsvariablen ist es sinnvoll, die im Prinzip willkürliche Zuweisung von Zahlen zu den beiden Kategorien durch Wertelabels zu dokumentieren, damit wir bei der Lektüre von Ergebnisausgaben nicht rätseln müssen, welches Geschlecht die Nummer Eins ist. Öffnen Sie daher mit einem Mausklick auf den Erweiterungsschalter ... in der markierten **Wertelabels**-Zelle die folgende Dialogbox:



Hier wird z.B. das weibliche Label folgendermaßen vereinbart:

- Tragen Sie den **Wert** 1 und das **Wertelabel** „Frau" ein
- drücken Sie dann auf den Schalter **Hinzufügen**.
  Weil in der Schaltflächen-Beschriftung **Hinzufügen** das **H** unterstrichen ist, können Sie den umständlichen Mausklick auf die Schaltfläche übrigens auch durch die Tastenkombination **<Alt>+<H>** ersetzen.

Einige Regeln für Variablen- und Wertelabel:

- Variablenlabel dürfen bis zu 120 Zeichen, Wertelabel bis zu 60 Zeichen lang sein. Allerdings erscheinen die Labels in der Ausgabe mancher SPSS-Prozeduren nicht in voller Länge.
- Umlaute und sonstige Sonderzeichen sind erlaubt.
- Groß/Kleinschreibung ist signifikant, wird also von SPSS respektiert.

Die Bedeutung der Wertelabel wird vielfach überschätzt. Beachten Sie bitte:

- Wertelabel dienen ausschließlich dazu, die Lesbarkeit der Ausgaben zu verbessern.
- Auf die Verarbeitung von Variablenausprägungen haben ggf. vorhandene Wertelabel keinerlei Konsequenzen, z.B. wird durch Wertelabel in keiner Weise die Menge der gültigen Werte einer Variablen festgelegt.

Statt die Breite der GESCHL-Spalte im Datenfenster über eine gut geschätzte **Spalten**-Angabe festzulegen, können Sie *bei aktiviertem Datenfenster-Registerblatt* **Datenansicht** auch folgendermaßen vorgehen: Setzen Sie den Mauszeiger auf den rechten Rand der Zelle mit dem Variablennamen, woraufhin der Zeiger eine neue Form und dementsprechend eine neue Funktion gewinnt. Nun lässt sich der rechten Rand der aktuellen Spalte verschieben: Linke Maustaste drücken, ziehen und an der gewünschten Position wider los lassen. Eine so festgelegte Spaltenbreite wird von SPSS als Variablenattribut übernommen.

3.2.2.4 Übung

Definieren Sie alle Variablen zu der ersten Seite unseres KFA-Fragebogens. Wie Sie nötigenfalls Variablen einfügen oder löschen können, erfahren Sie im nächsten Abschnitt.

3.2.3 Variablen einfügen, löschen oder verschieben

Bei der Variablendefinition kann sich durch Bedienungsfehler leicht die Notwendigkeit ergeben, Variablen einzufügen oder zu löschen.

3.2.3.1 Variablen einfügen

Wenn Sie z.B. nach FNR und GESCHL die Variable FB definiert und folglich die Variable GEBJ vergessen haben, können Sie das Missgeschick in der **Variablenansicht** folgendermaßen korrigieren:

- Markieren Sie die FB-Zeile durch einen Mausklick auf ihre Zeilennummer am linken Rand.
- Wählen Sie den Menübefehl:

  Daten > Variable einfügen

Alternativ kann auch einen rechten Mausklick auf die Beschriftung der FB-Zeile setzen und **Variable einfügen** aus dem Kontextmenü wählen.
In jedem Fall stellt SPSS anschließend vor FB eine neue Variable mit voreingestellten Attributen zur Verfügung, die nun beliebig angepasst werden können:



	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Werte	Spalten	Ausrichtung	Meßniveau
1	fnr	Numerisch	2	0		Kein	Kein	3	Rechts	Nominal
2	geschl	Numerisch	1	0	Geschlecht	{1, Frau}...	Kein	5	Rechts	Nominal
3	var00001	Numerisch	8	2		Kein	Kein	8	Rechts	Metrisch
4	fb	Numerisch	1	0		Kein	Kein	4	Rechts	Nominal

Auf analoge Weise lässt sich eine neue Variable auch in der **Datenansicht** einfügen:

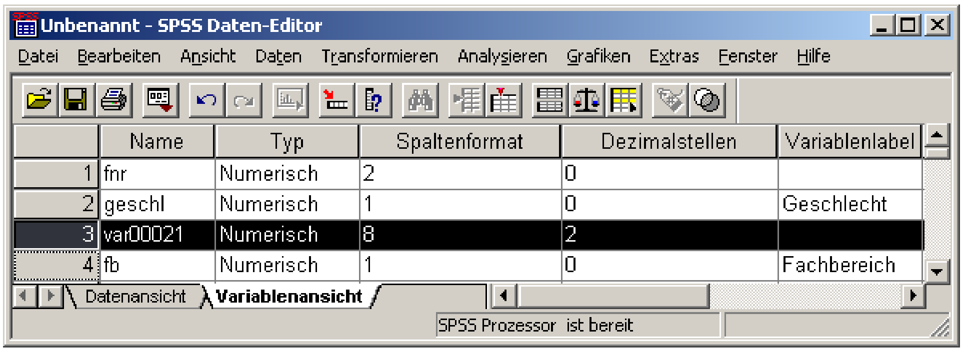
- Markieren Sie die FB-Spalte durch einen Mausklick auf ihre Beschriftung im Kopfbereich der Tabelle.
- Wählen Sie den Menübefehl:

  Daten > Variable einfügen

3.2.3.2 Variablen löschen

Gehen Sie in der **Variablenansicht** folgendermaßen vor, um eine Variable aus zu löschen:

- Markieren Sie die zugehörige Zeile durch einen Mausklick auf ihre Nummer am linken Tabellenrand, z.B.:



- Löschen Sie die markierte Variable mit der **<Entf>**-Taste oder mit dem Menübefehl

  Bearbeiten > Löschen

Auf analoge Weise lässt sich eine Variable auch in der **Datenansicht** löschen.

3.2.3.3 Variablen verschieben

Beim Verschieben einer Variablen sind Sie spätestens *dann* auf die Datenansicht angewiesen, wenn bereits Variablenausprägungen erfasst worden sind. Gehen Sie folgendermaßen vor, um eine Variable per *Drag & Drop* (ziehen und fallen lassen) zu verschieben:

- Markieren Sie die zu verschiebende Variable durch einen Mausklick auf ihren Namen im Tabellenkopf. Lassen Sie anschließend die Maustaste wieder los.
- Klicken Sie erneut auf den Namen der zu verschiebenden Variablen, und halten Sie dabei die Maustaste gedrückt.
- Bewegen Sie bei gedrückter Maustaste den Mauszeiger zum Ziel der Verschiebungsaktion. Der augenblicklich eingestellte Zielort wird von SPSS durch eine rote Linie gekennzeichnet.
- Wenn Sie die Maustaste los lassen, erscheint die Variable am neuen Ort.

Solange eine Variable noch keine Werte enthält, können Sie die beschriebene Verschiebungsaktion auf analoge Weise auch in der Variablenansicht des Datenfensters durchführen.

3.2.4 Attribute auf andere Variablen übertragen

3.2.4.1 Alle Attribute einer Variablen übernehmen

Für unsere 12 LOT-Fragen sollen natürlich alle Variablenattribute identisch sein. Erfreulicherweise müssen wir die identische Variablendefinition nicht 12 mal wiederholen, sondern können nach einer Deklaration der ersten Variablen die Attribute (mit Ausnahme des Namens) auf die anderen Variablen übertragen:

- Vergeben Sie für die nächsten 12 Variablen die Namen LOT01 bis LOT12[1].
- Legen Sie für LOT01 geeignete Attribute fest, z.B.:



[1] Leider müssen Sie tatsächlich alle 12 Namen einzeln schreiben. Später werden Sie sehen, wie man mit einem Dreizeiler in der SPSS-Kommandosprache beliebig viele Variablennamen automatisch eintragen lassen kann.

Das voreingestellte metrische Messniveau kann beibehalten werden, obwohl unsere fünfstufigen Variablen LOT01 bis LOT12 wohl eher grobschlächtige Indikatoren für die angenommene latente Variable Optimismus sind. In den geplanten Auswertungen werden wir nicht die Rohvariablen selbst, sondern eine daraus abgeleitete Mittelwertsvariable verwenden, für die ein approximativ metrisches Messniveau angenommen werden darf.

- Markieren Sie die komplette Variable LOT01 per Mausklick auf ihre Zeilennummer am linken Tabellenrand, und kopieren Sie alle Attribute mit **<Strg>+<C>** oder

  Bearbeiten > Kopieren

  in die Zwischenablage.
- Markieren Sie die Liste der Variablen von LOT02 bis LOT12, z.B. indem Sie Variable LOT02 markieren und dann bei gedrückter Umschalt-Taste zusätzlich die Variable LOT12 markieren.
- Übertragen Sie die in der Zwischenablage gespeicherten Attribute auf alle markierten Variablen mit **<Strg>+<V>** oder

  Bearbeiten > Einfügen

3.2.4.2 Einzelne Attribute einer Variablen übernehmen

Es ist auch möglich, ein *einzelnes* Attribut von einer Variablen auf mehrere andere zu übertragen:

- Quell-Attributzelle markieren
- Attribut mit **<Strg>+<C>** in die Zwischenablage kopieren
- Zu verändernde Attributzellen markieren
- Attribut mit **<Strg>+<V>** aus der Zwischenablage übernehmen

3.2.4.3 Übung

Definieren Sie die restlichen Variablen unserer KFA-Studie.

3.2.5 Sichern der Arbeitsdatei als SPSS-Datendatei

Wenn eine neu erstellte Arbeitsdatei über das Ende der Sitzung hinaus erhalten bleiben soll, muss sie explizit auf ein Festplatten-, Disketten- oder Netzlaufwerk gesichert werden. Dabei entsteht eine **SPSS-Datendatei**, früher auch als „SPSS-Systemdatei“ bezeichnet.
Nach dem *Öffnen* einer SPSS-Datendatei steht die dorthin gespeicherte Arbeitsdatei genau im gesicherten Zustand (also z.B. mit allen Variablendefinitionen) wieder zur Verfügung.
Zwar enthält Ihre Arbeitsdatei noch keine Daten, aber im Deklarationsteil stehen bereits wertvolle Informationen, deren Verlust recht schmerzlich wäre. Daher sollten Sie schon jetzt die temporäre Arbeitsdatei in eine permanente SPSS-Datendatei sichern, indem Sie den folgenden Menübefehl wählen:

Datei > Speichern unter...

In der erscheinenden Dialogbox können Sie für die zu erzeugende SPSS-Datendatei einen Namen, ein Verzeichnis, und ein Laufwerk angeben:



Wenn Sie die für SPSS-Datendateien vorgegebene Namenserweiterung ".sav" beibehalten, geht das spätere Öffnen der Datendatei besonders bequem.
Als Name für unsere Beispieldatei wird **kfar.sav** vorgeschlagen wird, verbunden mit der Versicherung, die Begründung für das „r“ im nächsten Abschnitt nachzuliefern. Wenn Sie an einem Pool-PC an der Universität Trier arbeiten, können Sie den automatisch erzeugten Ablageordner **U:\Eigene Dateien\SPSS**. benutzen.
Bestätigen Sie Ihre Eintragungen mit **OK**. Nach dem Sichern zeigt die Titelzeile des Datenfensters den Namen der nunmehr zugeordneten Datendatei, in unserem Fall also **kfar.sav**.
Beim Speichern einer Arbeitsdatei können auch alternative Dateiformate gewählt werden. Weil wir im weiteren Kursverlauf nicht mehr darauf eingehen, soll hier kurz erwähnt werden, dass Sie Ihre SPSS-Arbeitsdatei z.B. im XLS-Format des Tabellenkalkulationsprogramms Microsoft-EXCEL© abspeichern können.

Zum späteren Sichern in diese bereits eingestellte Datei dient der Befehl:

Datei > Speichern

Alternativ können Sie mit der Maus auf das Symbol  klicken oder die Tastenkombination **<Strg>+<S>** benutzen.

3.2.6 Rohdatendatei, Transformationsprogramm und Fertigdatendatei

Möglicherweise haben Sie sich beim Lesen des letzten Abschnitts gefragt, was das "r" im vorgeschlagenen Dateinamen **kfar.sav** bedeuten soll. Bei der Beantwortung dieser Frage sind leider einige Vorgriffe auf spätere Abschnitte nötig. Versuchen wir es trotzdem. Das "r" soll signalisieren, dass in dieser Datei die nach den Vorschriften des Kodierplans erfassten **R**ohdaten stehen. In **kfar.sav** sollen also ausschließlich folgende Arbeitsschritte einfließen:

- Variablendefinition gemäß Kodierplan
- Datenerfassung gemäß Kodierplan
- Nötigenfalls spätere Korrekturen von Erfassungsfehlern

Damit ist diese Datei für die meisten Auswertungsarbeiten noch nicht geeignet. Vor allem fehlen viele zu analysierende Variablen, die aus den Rohvariablen erst noch zu bilden sind, z.B. ein Optimismus-Testwert, der aus den 12 LOT-Fragen berechnet werden muss.

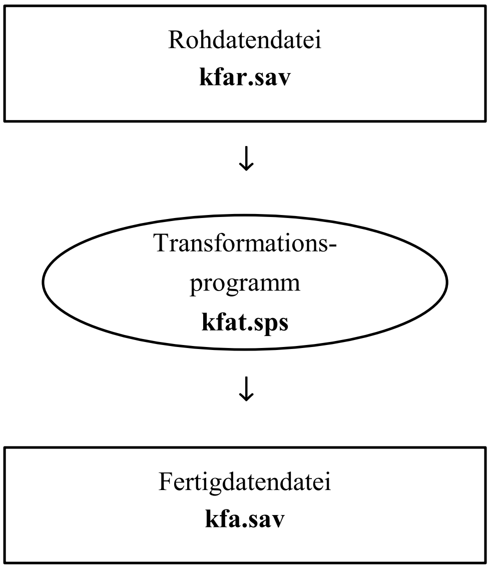
Aus der Rohdatendatei werden wir daher bald eine **Fertigdatendatei** herstellen, in die alle projektweit relevanten Variablenmodifikationen und -neuberechnungen einfließen sollen, so dass

sie eine bequeme Datenbasis für alle statistischen und graphischen Analysen darstellt. In fast jedem Projekt sind Variablenmodifikationen und -neuberechnungen in erheblichem Umfang erforderlich.

Profis modellieren dabei nicht „per Hand“ so lange an der Rohdatei herum, bis die Fertigdatei entstanden ist, sondern sie erstellen sich, z.B. durch Konservieren von bearbeiteten Dialogboxen, ein sogenanntes **SPSS-Programm** (siehe unten), das alle Transformationen erledigt und das bei Bedarf auch wiederholt ausgeführt werden kann.

Die zweistufige Projektdatenverwaltung mit Roh- und Fertigdatei verhindert in Kombination mit dem SPSS-Transformationsprogramm, dass bei jeder Änderung der Rohdaten die erwähnten Transformationen zur Fertigdatei „per Hand“ wiederholt werden müssen. Solche Änderungen der Rohdaten (z.B. durch Fehlerkorrekturen oder Stichprobenerweiterungen) sind eher die Regel als die Ausnahme.
Da die Kommandos des Transformationsprogramms auch mit Hilfe von korrespondierenden Dialogboxen erstellt werden können, erfordert die professionelle Vorgehensweise kaum Programmierkenntnisse.

Es wird also folgende Struktur für die Verwaltung der Projektdaten vorgeschlagen:



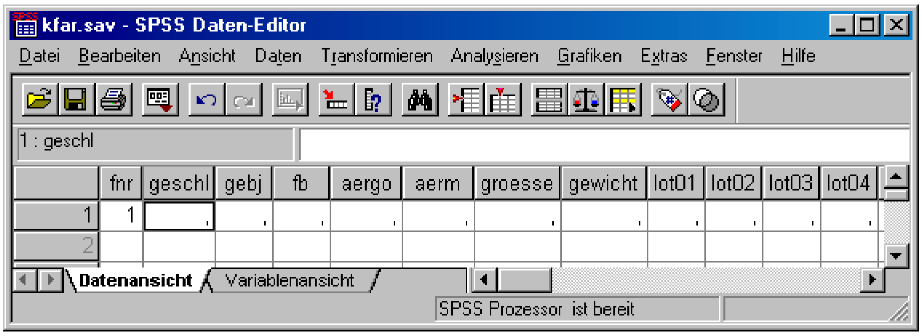
Die Erläuterungen in diesem Abschnitt werden vermutlich erst dann voll verständlich, wenn Sie sich mit Variablentransformationen und SPSS-Programmen auskennen.

Nach diesem Vorausblick wenden wir uns wieder der aktuellen Aufgabe zu: Wir tragen die erhobenen Daten in die Rohdatendatei **kfar.sav** ein.

3.2.7 Dateneingabe

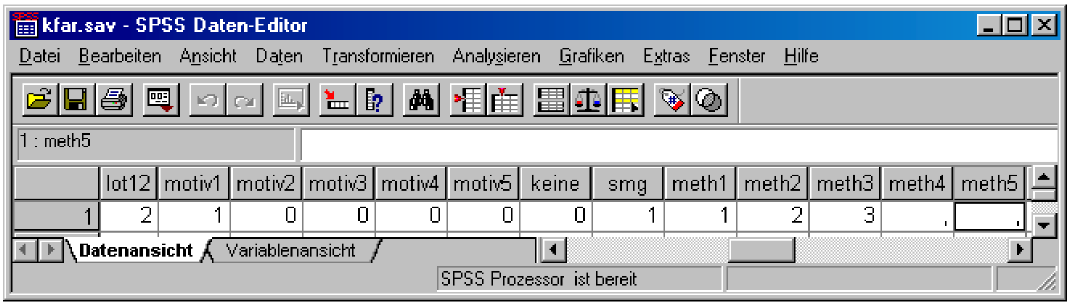
Wechseln Sie bei Bedarf zur Datenansicht, und geben Sie die Daten des ersten Falles ein:

- Aktivieren Sie nötigenfalls die Zelle zur ersten Variablen des ersten Falles und tippen Sie den zugehörigen Wert ein.
- Drücken Sie die Taste mit dem Rechtspfeil oder die **<Tab>**-Taste, um den eingetippten Wert zu quittieren und die Zellenmarkierung um eine Spalte nach *rechts* zu verschieben (zur nächsten Variablen):



Auch die **<Enter>**-Taste quittiert den eingetippten Wert, bewegt jedoch anschließend die Zellenmarkierung um eine Zeile nach *unten* (zum nächsten Fall), was in unserer jetzigen Lage weniger praktisch ist. Wenn Sie auf Abwege geraten sind, können Sie die Zellenmarkierung jederzeit per Mausklick neu positionieren.

- Sobald der Wert zur ersten Variablen eingetragen ist, erhält der bearbeitete Fall für die restlichen Variablen den Initialisierungswert SYSMIS (dargestellt durch ein Komma), was aber die weitere Dateneingabe nicht behindert.
  Tragen Sie die restlichen Werte des ersten Falles ein, jeweils quittiert mit der **<Tab>**-Taste.
  So sieht der vollständig erfasste erste Fall unserer Stichprobe im Datenfenster aus:



- Wenn Sie den Wert der letzten Variablen mit der **<Tab>**-Taste quittieren, setzt SPSS freundlicherweise die Zellenmarkierung gleich in die erste Datenzelle des nächsten Falles, so dass Sie die Dateneingabe unmittelbar fortsetzen können.

3.2.8 Daten korrigieren

3.2.8.1 Wert in einer Zelle ändern

Natürlich können die Eintragungen in einer Zelle jederzeit korrigiert werden:

- Wert ersetzen:
  - Zelle markieren
  - neuen Wert eintippen, wobei der alte überschrieben wird
  - neuen Wert quittieren, z.B. mit **<Enter>**
- Wert editieren:
  - Doppelklick auf die Zelle
  - Wert editieren
  - neuen Wert quittieren, z.B. mit **<Enter>**

3.2.8.2 Einen Fall einfügen

Gehen Sie folgendermaßen vor, um einen Fall, d.h. eine Zeile der Datenmatrix, einzufügen:

- Klicken Sie die Beschriftung desjenigen Falles an, *vor* dem ein neuer Fall eingefügt werden soll. Daraufhin wird die gesamte angeklickte Zeile markiert.
- Wählen Sie den Menübefehl

  Daten > Fall einfügen

Der neue Fall erhält bei allen Variablen den Wert System-Missing. Diese Initialisierungswerte können dann natürlich beliebig überschrieben werden.

3.2.8.3 Einen Fall löschen

Gehen Sie folgendermaßen vor, um einen Fall, d.h. eine Zeile der Datenmatrix, komplett zu löschen:

- Klicken Sie die Beschriftung des überflüssigen Falles an. Daraufhin wird die gesamte Zeile markiert.
- Wählen Sie den Menübefehl

  Bearbeiten > Löschen

  oder die **<Entf>**-Taste.

3.2.8.4 Einen Fall verschieben

Gehen Sie folgendermaßen vor, um einen Fall zu per *Drag & Drop* (ziehen und fallen lassen) zu verschieben:

- Klicken Sie die Fall-Beschriftung an. Daraufhin wird die gesamte Zeile markiert. Lassen Sie anschließend die Maustaste wieder los.
- Klicken Sie erneut auf die Fall-Beschriftung, und halten Sie dabei die Maustaste gedrückt.
- Bewegen Sie bei gedrückter Maustaste den Mauszeiger zum Ziel der Verschiebungsaktion. Der augenblicklich eingestellte Zielort wird von SPSS durch eine rote Linie gekennzeichnet.
- Wenn Sie die Maustaste los lassen, erscheint der Fall am neuen Ort.

3.2.9 Weitere Möglichkeiten des Dateneditors

Über die beschriebenen Methoden hinaus bietet der Dateneditor u.a. die Möglichkeit, beliebige rechteckige Segmente der Datenmatrix auszuschneiden, zu kopieren und einzufügen. Wer derartige, relativ fehleranfällige Umordnungsmaßnahmen vornehmen muss, wird gelegentlich von der Möglichkeit profitieren, mit:

Bearbeiten > Rückgängig

die letzte Änderung rückgängig machen zu können.

In Abschnitt 5.6 wird beschrieben, wie Sie im Datenfenster nach Variablenausprägungen suchen können.

Eine Erklärung aller Funktionen des Dateneditors findet sich im Online-Hilfesystem:

Hilfe > Themen > Inhalt > Datenverwaltung > Arbeiten mit dem Daten-Editor

3.2.10 Übung

Für die Teilnehmer(innen) des realen SPSS-Kurses steht nun die etwas mühselige Erfassung der erhobenen Daten an. Geben Sie alle Fälle ein, und sichern Sie (auch zwischendurch) in die zugeordnete Datendatei, z.B. **U:\Eigene Dateien\SPSS\kfar.sav**.

Wer dem Vorschlag in diesem Manuskript folgend zur Erfassung der Antworten auf die offene Frage im Fragebogenteil 4b) ein dynamisches und sparsames Set aus kategorialen Variablen vorgesehen hat (z.B. METH1 bis METH5), der muss nicht nur mechanisch Daten eintippen, sondern auch gelegentlich mit Kreativität und Ordnungssinn neue Methoden-Kategorien definieren und dokumentieren.
Beim Erfassen der Daten, die in diesem Manuskript analysiert werden, wurden folgende Kategorien gebildet und mit einem numerischen Code versehen:

Kategorie	Code
Faktorenanalyse	1
Regressionsanalyse	2
Korrelationsanalyse	3
Varianzanalyse	4
Strukturgleichungsanalyse	5
Clusteranalyse	6
Diskriminanzanalyse	7
Logistische Regression	8
Conjoint-Analyse	9

Diese Tabelle vervollständigt unseren Kodierplan in Abschnitt 1.4.3.5.

Es bietet sich an, die Definition der Variablen METH1 bis METH5 durch entsprechende Wertelabels zu vervollständigen (vgl. Abschnitt 3.2.2.3), damit die späteren Ergebnisausgaben leichter zu verstehen sind. Die für Leser(innen) im Selbststudium via WWW (siehe Vorwort) zur Verfügung gestellte KFA-Rohdatendatei **kfar.sav** enthält solche Label-Definitionen.

4 Datenprüfung

Bei der manuellen Datenerfassung sind Fehler praktisch unvermeidbar. Manche von ihnen sind als Verstöße gegen allgemeine Gültigkeitsregeln relativ leicht aufzuspüren:

Beispiel: Wenn bei der Variablen GESCHL nur die Werte 1 (für Frauen) und 2 (für Männer) erlaubt sind, dann ist z.B. der Wert 3 sofort als falsch erkennbar.

Weit schwieriger zu entdecken sind Fehler, die keine allgemeine Gültigkeitsregel verletzen:

Beispiel: Wenn unter der oben angegebenen GESCHL-Kodierungsvorschrift für den Untersuchungsteilnehmer Kurt Müller versehentlich der Wert 1 eingegeben wurde, dann kann dieser Fehler nur durch aufwendige Handarbeit gefunden werden.

Welcher Aufwand bei der Datenprüfung erforderlich bzw. sinnvoll ist, hängt wesentlich davon ab, wie die Daten erfasst worden sind (vgl. Abschnitt 3.1).
"Nobody is perfect" gilt übrigens nicht nur für Menschen, sondern auch für Maschinen. Daher sollte man vorsichtshalber auch bei Verwendung einer automatischen Erfassungsmethode stichprobenartig die Datenintegrität überprüfen.

Nach der Erfassung per Texteditor ist die Menge potentieller Fehler besonders groß. Deshalb wurde oben nachdrücklich von dieser veralteten Erfassungsmethode abgeraten. Konsequenterweise gehen wir auch im Abschnitt über Datenprüfung nicht auf die speziellen Probleme ein, die nach dem Erfassen per Texteditor auftreten können.

4.1 Überprüfung von Gültigkeitsregeln

Wir beschränken uns auf die Suche nach **ungültigen Werten** (wie z.B. eine negative Gewichtsangabe), wenngleich damit nicht alle Möglichkeiten zum Aufspüren von verletzten Gültigkeitsregeln ausgereizt werden. Weitere Hinweise sind z.B. bei Bauer (1986, S. 21ff) zu finden.

Bei der Erfassung per Datenbankprogramm mit Plausibilitätskontrolle werden unzulässige Werte zurückgewiesen und folglich von der Datendatei fern gehalten. Bei der Erfassung mit dem SPSS-Dateneditor findet eine derartige Eingangskontrolle nicht statt. Eine so entstandene Datei muss daher systematisch nach Daten außerhalb des zulässigen Bereichs durchsucht werden. Dies kann allerdings ohne großen Zusatzaufwand im Rahmen der aus wissenschaftlichen Gründen ohnehin erforderlichen univariaten Verteilungsanalyse geschehen (siehe Abschnitt 5).

4.2 Überprüfung von Einzelwerten

Fehler, die gegen keine Gültigkeitsregel verstoßen, lassen sich nur mit Fleißarbeit entdecken, wobei z.B. folgende Vorgehensweisen möglich sind:

- Man vergleicht die erfassten Daten Wert für Wert mit den schriftlichen Unterlagen.
- Manche Datenbankprogramme versuchen, die Erfasser durch Kontrollen und Sanktionen zu sorgfältiger Arbeit zu motivieren: INPUT II erlaubt z.B. die Festlegung einer Kontrollwahrscheinlichkeit, mit der ein Erfasser einen Teil des letzten Datensatzes nochmals eingeben muss. Bei Erfolg sinkt die Kontrollwahrscheinlichkeit, bei Misserfolg werden die diskrepanten Daten präsentiert, und die Kontrollwahrscheinlichkeit steigt.

Eine aufwendige Prüfmethode ist *bei kleinen Stichproben* durchaus empfehlenswert, denn:

- Hier können sich Erfassungsfehler besonders stark auswirken:
    - Sie können Schätzergebnisse erheblich verzerren. In kleinen Stichproben darf man nicht darauf vertrauen, dass sich verschiedene Fehler ausmitteln.
    - Sie erhöhen die Fehlervarianz und verringern damit die Power der Hypothesentests (hoher β-Fehler), was sich vor allem in kleinen Stichproben mit ihren chronischen Power-Problemen negativ bemerkbar macht.
- Der Zeitaufwand ist erträglich.

5 Univariate Verteilungs- und Fehleranalysen

In diesem Abschnitt werden Sie erfahren, wie schnell und bequem mit SPSS numerische und graphische Analysen durchgeführt werden können. Wir werden unsere Daten mit Hilfe deskriptiver Auswertungsmethoden sorgfältig auf Erfassungsfehler untersuchen. Dabei schlagen wir zwei Fliegen mit einer Klappe, denn eine sorgfältige Verteilungsanalyse aller Variablen gehört ohnehin zur Pflicht bei jeder empirischen Studie.
In manchen Projekten wird sich die Forschungsarbeit sogar auf die Beschreibung der Verteilungen von wichtigen Variablen beschränken (z.B. in der Meinungsforschung). Meist sind aber auch multivariate Zusammenhangsanalysen von Interesse.

5.1 *Öffnen einer SPSS-Datendatei*


Ich gehe davon aus, dass Sie nach der anstrengenden Datenerfassung eine Pause eingelegt und SPSS verlassen haben. Damit können wir also jetzt offiziell die Fortsetzung einer unterbrochenen Projektarbeit üben. Starten Sie SPSS, und öffnen Sie Ihre vorhandene Rohdatendatei **kfar.sav**, entweder mit Hilfe des Startassistenten oder über den Menübefehl

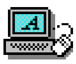
Datei > Zuletzt verwendete Daten

Beim Öffnen einer Datendatei legt SPSS eine neue (temporäre) Arbeitsdatei an und kopiert die eingelesenen Daten samt Variablendeklarationen dorthin. Alle Veränderungen, die Sie in der Datenmatrix oder im Deklarationsteil vornehmen, wirken sich zunächst nur auf die temporäre Arbeitsdatei aus. Gegebenenfalls müssen Sie also diese Änderungen über den Menübefehl

Datei > Speichern

in die permanente SPSS-Datendatei **kfar.sav** übernehmen.

5.2 *Statistische Auswertungen durchführen: Häufigkeitsanalyse zur Prüfung der Variablen FNR*


Da wir unsere Daten mit dem SPSS-Dateneditor erfasst haben, der keine Plausibilitätskontrolle bei der Eingabe vornimmt, müssen wir nach den Überlegungen aus Abschnitt 4 systematisch nach unzulässigen Werten suchen. Die meisten der dazu erforderlichen deskriptiven Datenanalysen wären im Rahmen der routinemäßigen Verteilungsuntersuchung ohnehin fällig gewesen.
Der erste Test dient allerdings ausschließlich zur Datenprüfung, weil dabei die Fallidentifikations-Variable FNR untersucht wird. Weil die Manuskript-Stichprobe den Umfang N = 31 hat, und es keinen Grund für eine „lückenhafte“ Nummerierung gab, müssen nach fehlerfreier Erfassung bei dieser Variablen die Werte 1, .., 31 jeweils genau einmal auftreten. Daraus ergeben sich einige notwendige Bedingungen, die sich leicht nachprüfen lassen:

- In der Stichprobe müssen bei der Variablen FNR 31 gültige Werte vorliegen. (MD-Indikatoren sind hier nicht erlaubt.)
- Der kleinste Wert muss gleich 1, und der größte Wert muss gleich 31 sein.
- Jeder Wert darf höchstens einmal auftreten, d.h. der Stichproben-Modus muss die Häufigkeit 1 haben.

Zur Überprüfung der Bedingungen lassen wir in einer *Häufigkeitsanalyse* für die Variable FNR folgende Statistiken berechnen: Anzahl valider Fälle, Minimum, Maximum und Modus.
Mit dem Menübefehl

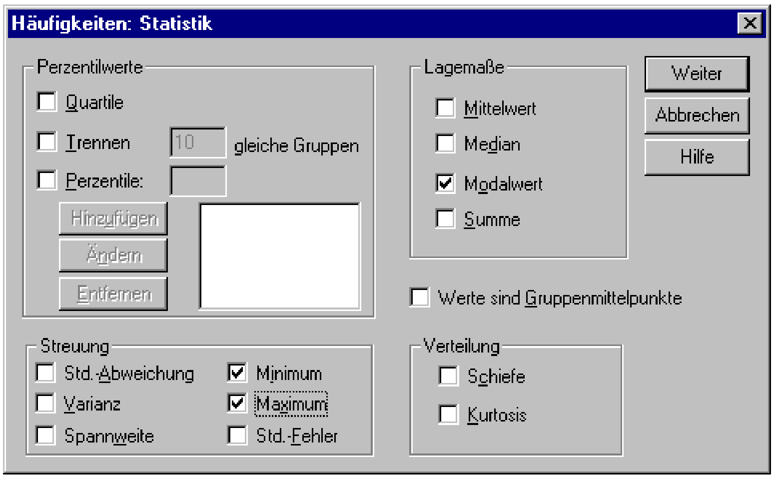
Analysieren > Deskriptive Statistik > Häufigkeiten...

erhalten wir die folgende Dialogbox zur Anforderung von Häufigkeitsanalysen:



Zur bequemen Spezifikation der im aktuellen Prozeduraufruf zu analysierenden Variablen dienen die beiden Variablen-Auswahlbereiche. Links stehen alle Variablen der Arbeitsdatei, die derzeit für die Analyse *nicht* ausgewählt sind („Anwärterliste"). Rechts daneben, im Bereich **Variable(n)**, stehen die Ausgewählten („Teilnehmerliste"). Dazwischen befindet sich ein Transportschalter, mit dem sich links markierte Variablen nach rechts und rechts markierte Variablen nach links verschieben lassen. Markieren Sie also links die Fallnummern-Variable FNR und drücken Sie auf den Transportknopf.

Zur Auswahl der gewünschten **Statistiken** müssen Sie die zuständige Subdialogbox durch Knopfdruck aktivieren. Um eine der hier aufgelisteten Möglichkeiten zu wählen, ist das zugehörige Kontrollkästchen zu markieren:



Quittieren Sie nun die Subdialogbox mit **Weiter** und die Hauptdialogbox mit **OK**. Daraufhin führt SPSS die Berechnungen aus und präsentiert die Ergebnisse im Ausgabefenster (SPSS Viewer), das sich in den Vordergrund drängt.

Bei Anforderung einer Häufigkeitsanalyse produziert SPSS per Voreinstellung eine Häufigkeitstabelle, die für jeden aufgetretenen Wert eine Zeile mit folgenden Angaben enthält:

- Absolute Häufigkeit
- Relative Häufigkeit in Prozent
- Prozentualer Anteil an den validen Werten (ohne MD-deklarierte Werte)
- kumulativer Prozentanteil (Anteil valider Werte, die nicht größer sind)

Außerdem berichtet SPSS unaufgefordert, wie viele Fälle einen validen Wert bzw. einen MD-deklarierten Wert haben. Weitere Leistungen müssen explizit angefordert werden.

Obige Dialogbox liefert folgende Statistiken:

Statistiken

Fallnummer

N	Gültig	31
	Fehlend	0
Modus		1[a]
Minimum		1
Maximum		31

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

Indizien für Erfassungsfehler finden sich nicht: Alle 31 Personen haben einen validen Wert, das Minimum ist 1, das Maximum ist 31.
Laut Häufigkeitstabelle (hier verkürzt wiedergegeben) hat der (natürlich nicht eindeutige) Modalwert die Häufigkeit 1:

Fallnummer

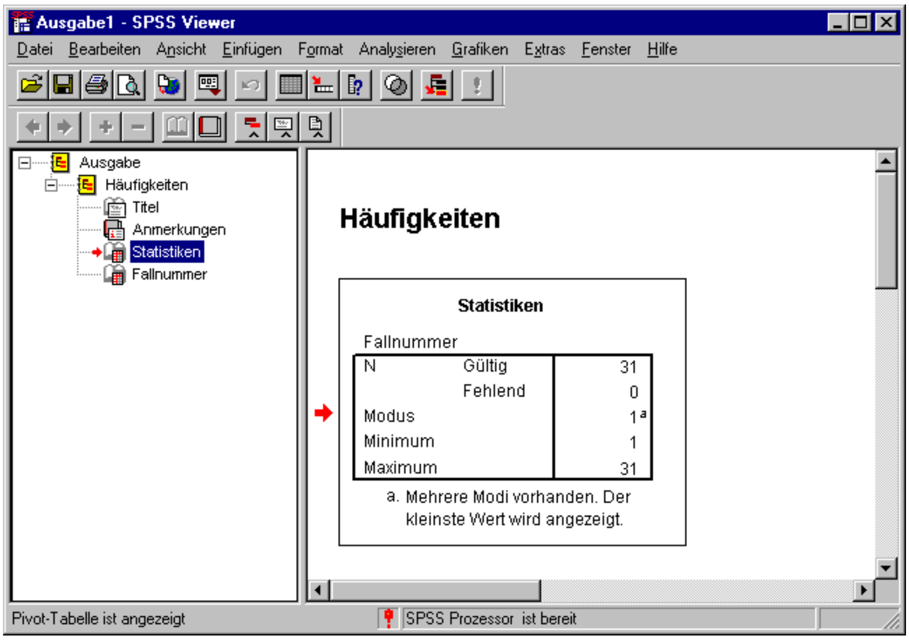
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1	1	3,2	3,2	3,2
	2	1	3,2	3,2	6,5
	3	1	3,2	3,2	9,7
	4	1	3,2	3,2	12,9
	,	,	,	,	,
	,	,	,	,	,
	,	,	,	,	,
	30	1	3,2	3,2	96,8
	31	1	3,2	3,2	100,0
	Gesamt	31	100,0	100,0	

Wir haben uns bei der FNR-Prüfung auf einige *notwendige* Bedingungen beschränkt, weil momentan nur elementare SPSS-Operationen benutzt werden sollen. Ein Perfektionismus ist bei dieser von uns selbst erzeugten Variablen ohnehin nicht erforderlich.

Die obigen SPSS-Ausgaben wurden übrigens aus dem Ausgabefenster via Windows-Zwischenablage in Microsoft Word© übertragen. Mit dieser Form des Datenaustauschs und mit anderen Möglichkeiten beim Arbeiten mit dem Ausgabefenster (Viewer) beschäftigen wir uns im nächsten Abschnitt.

5.3 Arbeiten mit dem Ausgabefenster (Teil I)

In seiner voreingestellten Variante ist das SPSS-Ausgabefenster, das auch als **Viewer** bezeichnet wird, zweigeteilt in die Gliederungsansicht am linken Rand und den eigentlichen Inhaltsbereich:



So soll ein schnelles Navigieren zwischen den verschiedenen Ausgabebestandteilen ermöglicht werden.

Die Aufteilung des verfügbaren Platzes auf die beiden Teile des Viewers kann per Maus beliebig verändert werden: Trennlinie anklicken und bei gedrückter Maustaste horizontal verschieben.

Wesentliche Bestandteile des Inhaltsbereichs sind Pivot-Tabellen, Grafiken und Textausgaben. Zu ihrer Nachbearbeitung steht jeweils ein spezieller Editor zur Verfügung, der per Doppelklick auf das Objekt gestartet wird. Außerdem können in einem Viewer-Dokument noch protokollierte SPSS-Anweisungen, Warnungen, Anmerkungen und Titelzeilen auftreten.

5.3.1 Arbeiten mit dem Navigationsbereich

Die meisten der anschließend beschriebenen Aktionen im Navigationsbereich wirken sich analog auch auf den Inhaltsbereich aus.

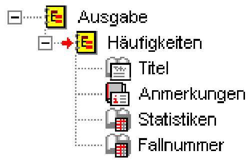
Mit **Bearbeiten > Rückgängig** bzw. **<Strg>+<Z>** kann in vielen Situationen die letzte Modifikation rückgängig gemacht werden.

5.3.1.1 Fokus positionieren

Ein kleiner roter Pfeil deutet im Gliederungs- bzw. Navigationsbereich auf die Bezeichnung derjenigen Ausgabe, die im Inhaltsbereich gerade privilegiert dargestellt wird. Per Mausklick auf eine andere Ausgabenbeschriftung kann dieser Fokus beliebig verschoben werden.

5.3.1.2 Ausgabeblöcke bzw. Teilausgaben aus- oder einblenden

Ein *Block* mit zusammengehörigen Ausgaben (in der Regel entstanden aus einer Analyseanforderung) wird ...

- ausgeblendet: per Mausklick auf das Minus-Zeichen neben Block-Symbol und Block-Beschriftung oder per Doppelklick auf das Block-Symbol.
  Beispiel:
  

- eingeblendet: per Mausklick auf das Plus-Zeichen neben Block-Symbol und Block-Beschriftung oder per Doppelklick auf das Block-Symbol.
  Beispiel:
  

Eine *Teil*ausgabe innerhalb eines Blockes wird per Doppelklick auf das zugehörige Buchsymbol aus- bzw. eingeblendet. Das Buchsymbol erscheint dementsprechend zugeklappt (im Beispiel: Anmerkungen) oder aufgeklappt (im Beispiel: Statistiken).

5.3.1.3 Ausgabeblöcke oder -teile markieren

Im Navigationsbereich können Sie auf windows-übliche Weise Ausgabeblöcke und/oder Teilausgaben markieren:

- Einen Ausgabeblock: Per Mausklick auf das Block-Symbol oder auf die Beschriftung
- Eine Teilausgabe: Per Mausklick auf das Buchsymbol oder auf die Beschriftung
- Mehrere Blöcke bzw. Teile: Per **<⇑>**-Mausklick bzw. **<Strg>**-Mausklick

Sie können markierte Blöcke bzw. Teilausgaben z.B. mit der **<Entf>**-Taste löschen oder mit dem Menübefehl **Bearbeiten > Objekte kopieren** in die Windows-Zwischenablage befördern.

5.3.2 Viewer-Dokumente drucken

Über den Standardbefehl **Datei > Drucken** können Sie alle angezeigten oder alle markierten Ausgabebestandteile drucken, wobei ein grafikfähiger Drucker benötigt wird. Gut geeignet sind z.B. die in vielen Poolräumen der Universität verfügbaren Postscript-Laserdrucker. Da hierbei verschiedene Typen zum Einsatz kommen, müssen Sie im Kombifeld **Name** der **Drucken**-Dialogbox den Typ des gewünschten Druckers einstellen.
Später lernen Sie eine Möglichkeit kennen, die per Voreinstellung grafikorientierten SPSS-Ausgaben durch einfachere Varianten zu ersetzen, die mit jedem Textdrucker ausgegeben werden können.

5.3.3 Ausgaben sichern und öffnen

Zum Speichern eines Viewer-Dokumentes dienen die Menübefehle **Datei > Speichern unter** bzw. **Datei > Speichern**. Dabei entstehen Viewer-Dateien, die üblicherweise durch die Namenserweiterung „.spo" gekennzeichnet werden. SPSS-Ausgaben sollten z.B. *dann* in elektronischer Form gespeichert werden, wenn sie (auszugsweise) in Dokumente anderer Programme eingegangen sind, z.B. in Word-Dateien. Mit SPSS ist eine nachträgliche Modifikation dieser Ausgaben leicht möglich, mit Fremdprogrammen aber kaum.

Zum Öffnen eines Viewer-Dokumentes mit den Befehlen **Datei > Öffnen > Ausgabe** oder **Datei > Zuletzt geöffnete Dateien** gibt es nichts Ungewöhnliches zu berichten.

5.3.4 Objekte via Zwischenablage in andere Anwendungen übertragen

SPSS kann markierte Ausgabe-Objekte (auch mehrere gleichzeitig) in verschiedenen Formaten in die Zwischenablage übertragen, aus der sie von anderen Anwendungen übernommen werden können. Bei der Übertragung im **Grafikformat** landen alle Objekte (z.B. Grafiken, Pivot-Tabellen) in der Zielanwendung genau so, wie sie im Viewer ausgesehen haben. Um dies zu erreichen, sollten Sie im SPSS-Viewer die markierten Objekte mit der Tastenkombination **<Strg>+<K>** oder mit dem Menübefehl

Bearbeiten > Objekte kopieren

in die Zwischenablage befördern. Zum Einfügen in der Zielanwendung können Sie den Menübefehl **Bearbeiten > Einfügen** bzw. die Tastenkombination **<Strg>+<V>** verwenden. So wurden z.B. die in Abschnitt 5.2 wiedergegebenen Tabellen übertragen.

In der Zielanwendung können Sie die Größe und Platzierung der übertragenen Objekte neu festlegen. Zumindest für die Zielanwendung Microsoft Word© gilt jedoch, dass eine weitere Nachbearbeitung der übertragenen Objekte leider **nicht** möglich ist. Wenn Sie es doch versuchen und per Doppelklick auf das eingefügte Objekt den Word-internen Grafiker aktivieren, wird dieser die Bestandteile der Grafik oder Tabelle auf unschöne Weise neu anordnen.

Wenn bei Tabellen eine freie Formatierungsmöglichkeit in der Zielanwendung gefragt ist, sollte die Übertragung im RTF-Format gewählt werden (siehe Abschnitt 5.7.3).

5.3.5 Übungen

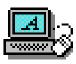
1) Markieren Sie den Ausgabeblock mit der Häufigkeitsanalyse, und löschen Sie ihn mit der **<Entf>**-Taste.

2) Steigen Sie erneut in die Dialogbox zur Häufigkeitsanalyse ein. Statt den zugehörigen Menübefehl zu wiederholen, können Sie einfacher mit dem Symbol  eine Liste der zuletzt benutzten Dialogboxen aufrufen und daraus per Mausklick den Eintrag **Häufigkeiten** wählen. Die Dialogbox ist noch im selben Zustand, den Sie eben verlassen haben. Dies gilt selbstverständlich generell in SPSS, so dass Sie bei der sukzessiven Modifikation einer Anforderung innerhalb einer Sitzung jeweils auf dem letzten Stand weitermachen können.

3) Schalten Sie die Häufigkeitstabelle über das zugehörige Kontrollkästchen aus, und lassen Sie die Häufigkeitsanalyse erneut ausführen.

4) Wie hätte man den jetzt erreichten Zustand des Ausgabefensters auch einfacher herstellen können?

5.4 Grafische Darstellungen in Statistik-Dialogboxen anfordern: Häufigkeits- bzw. Fehleranalyse für die Variablen GESCHL und FB



Nun wollen wir weitere Variablen untersuchen und dabei auch graphische Verteilungsdarstellungen verwenden. Dazu rufen wir erneut die Dialogbox zur Häufigkeitsanalyse auf und beseitigen alle alten Festlegungen (auch in den Subdialogboxen) mit dem Schalter **Zurücksetzen**.

Dann transportieren wir nacheinander die Variablen GESCHL und FB aus der Anwärterliste (links) in die Teilnehmerliste (rechts).

Anschließend begeben wir uns in die Subdialogbox **Diagramme** und entscheiden uns im Optionenfeld **Diagrammtyp** für Balkendiagramme, weil die Merkmale Geschlecht und Fachbereich nominalskaliert sind. Wer nicht mehr genau weiß, wozu man Balkendiagramme und Histogramme verwendet, kann sich mit der kontextsensitiven **Hilfe** Aufklärung verschaffen.

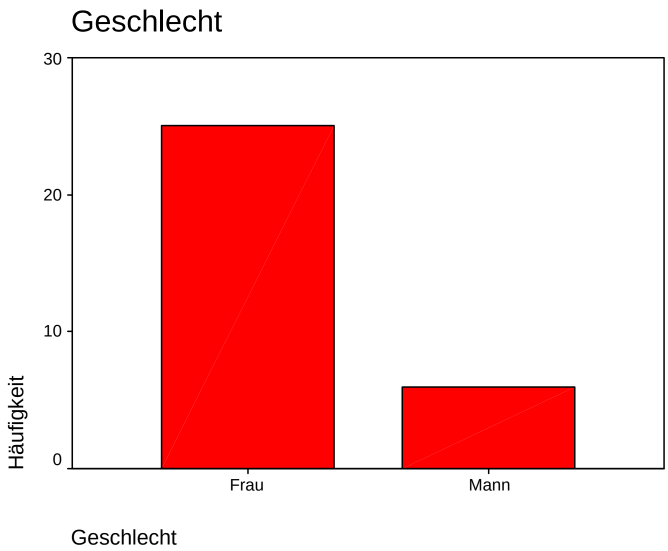
Jetzt starten wir die neue Analyse, indem wir die Subdialogbox mit **Weiter** und die Hauptdialogbox mit **OK** quittieren.

Im Viewer erhalten wir für die Variable GESCHL die Häufigkeitstabelle

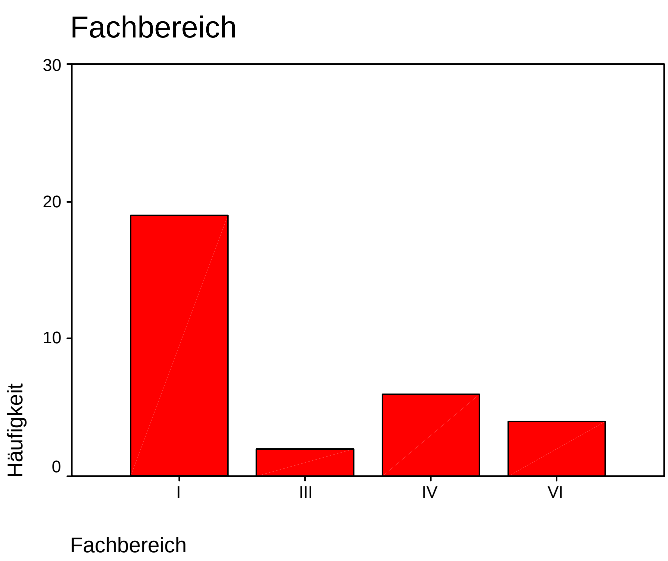
Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Frau	25	80,6	80,6	80,6
	Mann	6	19,4	19,4	100,0
	Gesamt	31	100,0	100,0	

und das folgende Balkendiagramm:



Zunächst beobachten wir, dass bei der Variablen GESCHL kein unzulässiger Wert vorliegt.
Bei der Geschlechtsverteilung stellen wir einen erstaunlich hohen Frauenanteil fest, der als wesentliches Merkmal unserer Stichprobe berichtet werden muss. Bei potentiell geschlechtsabhängigen Ergebnissen müssen wir besonders vorsichtig interpretieren und generalisieren.
Erste Hinweise zur Ursache der hohen Frauenquote liefert die empirische Verteilung der Fachbereichs-Variablen:



Wir sehen, dass in dem SPSS-Kurs, der die Manuskript-Daten geliefert hat, der Fachbereich I sehr stark vertreten war, was mit dem Kurstermin zusammenhängen mag. Im Fachbereich I der Universität Trier (Fächer: Philosophie, Pädagogik, Psychologie) ist aber bekanntermaßen der Frauenanteil sehr hoch.
Obige Abbildungen wurden übrigens mit der in Abschnitt 5.3.4 beschriebenen Methode vom SPSS-Viewer in WinWord übertragen.

Der aktuelle Abschnitt sollte nur einen ersten Eindruck von den Grafikmöglichkeiten des SPSS-Systems vermittelt. Wir haben eine integrierte Grafik-Option der Dialogbox zur Häufigkeitsanalyse benutzt. Die meisten grafischen Darstellungsmöglichkeiten bietet SPSS über das Hauptmenü **Grafiken** an, mit dessen Optionen wir uns später befassen werden.

5.5 Häufigkeits- bzw. Fehleranalysen für die restlichen Projektvariablen

5.5.1 Übung

Mittlerweile verfügen Sie über genügend SPSS-Kenntnisse, um die restlichen Häufigkeits- bzw. Fehleranalysen zu unserem Projekt selbständig durchführen zu können:

1) Die Merkmale Geburtsjahr, Größe, Gewicht und die beiden Ärgermaße können näherungsweise als metrisch angesehen werden. Lassen Sie sich daher für die zugehörigen Variablen ausgeben:
   - keine Häufigkeitstabellen
     Das für Tabellen zuständige Kontrollkästchen in der Dialogbox **Häufigkeiten** ist per Voreinstellung angekreuzt. Sie müssen also die Markierung durch Anklicken beseitigen.
   - Histogramme mit eingezeichneter Normalverteilungsdichte
   - folgende Statistiken: Mittelwert, Median, Modalwert, Standardabweichung, Varianz, Minimum, Maximum, Schiefe, Kurtosis (Exzeß)

   Hinweise zu einigen Statistiken:

   Schiefe

   Bei symmetrischen Variablen ist die Schiefe gerade gleich 0. Sie wird positiv bei linkssteil (bzw. rechtsschief) verteilten Variablen, wenn also die Verteilungsmasse am linken Rand konzentriert ist, und negativ bei rechtssteil (bzw. linksschief) verteilten Variablen.
   Zur Schiefe wird auch der zugehörige Standardfehler ausgegeben, mit dessen Hilfe wir Tests zur Populations-Schiefe veranstalten können. Diese sind allerdings nur approximativ gültig und vor allem in kleineren Stichproben mit Vorsicht zu genießen. Ihr Vorzug gegenüber später vorzustellenden Verteilungs-Anpassungstests besteht darin, dass sie gezielt auf Verletzungen der Verteilungs-*Symmetrie* ansprechen.
   Bei einem α–Fehlerrisiko von 5 % ist die *zwei*seitige Nullhypothese, dass die Schiefe in der Population gerade gleich 0 sei, zu verwerfen, falls:

   $$\frac{|\text{Schiefe}|}{\text{SF(Schiefe)}} > 1{,}96$$

   Beim selben α–Niveau ist die *ein*seitige Nullhypothese, dass die Populations-Schiefe größer oder gleich 0 sei, zu verwerfen, falls:

   $$\frac{\text{Schiefe}}{\text{SF(Schiefe)}} < -1{,}65$$

Analog lässt sich natürlich auch die einseitige Nullhypothese mit umgekehrtem Vorzeichen prüfen.

Kurtosis (Exzeß)

Der Exzeß (synonym: Kurtosis, Breitgipfligkeit, Wölbung) ist bei normalverteilten Variablen gleich 0. Er wird negativ bei breiteren und positiv bei schlankeren Verteilungen.
Mit Hilfe des Standardfehlers können analog zum Vorgehen bei der Schiefe-Statistik (siehe oben) „quick-and-dirty-Tests“ zum Exzeß in der Population durchgeführt werden.

2) Lassen Sie sich für die LOT-Variablen ausgeben:
   - Häufigkeitstabellen
   - keine Grafiken
   - folgende Statistiken: Mittelwert, Median, Modalwert, Standardabweichung, Varianz, Minimum, Maximum

3) Lassen Sie sich für die Variablen MOTIV1 bis MOTIV5, KEINE, SMG und METH1 bis METH5 ausgeben:
   - Häufigkeitstabellen
   - keine Grafiken
   - keine Statistiken

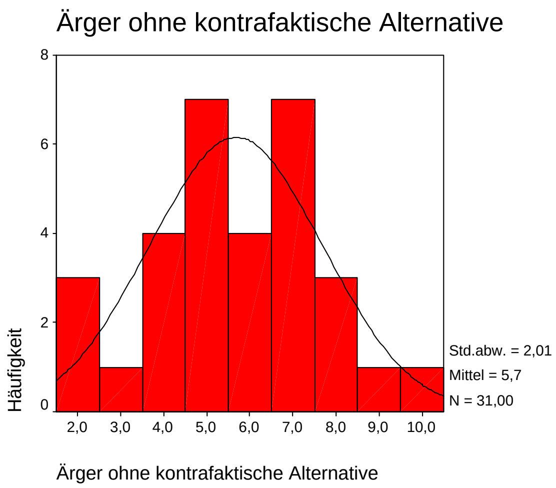
4) Prüfen Sie für alle Variablen nach, ob unzulässige Werte vorliegen.

5) Untersuchen Sie bei den metrischen Variablen GROESSE, GEWICHT, AERGO und AERGM zusätzlich, ob diese annähernd normal verteilt sind. Beziehen Sie in Ihr Urteil die Statistiken Schiefe und Kurtosis sowie deren Standardfehler ein.
   Die Vergleiche mit der Normalverteilung erfolgen hier aus purem Interesse an den Verteilungen der betrachteten Variablen, ohne dabei bereits an die Verteilungsvoraussetzungen irgendwelcher Testverfahren zu denken. Diese Voraussetzungen beziehen sich ohnehin häufig nicht auf die momentan von uns analysierten univariaten Verteilungen, sondern z.B. auf bedingte Verteilungen bzw. auf die Verteilungen der Residuen eines bestimmten statistischen Modells. Nähere Aussagen sind nur im Zusammenhang mit konkreten Testverfahren möglich. Zur Prüfung von Normalverteilungsvoraussetzungen setzt man zudem an Stelle der beiden oben beschriebenen „quick-and-dirty-Tests“ effizientere Verfahren ein (z.B. den Shapiro-Wilk - Test). Ein Vorzug der oben beschriebenen Tests besteht allerdings darin, dass sie einzelne Verteilungsaspekte gezielt ansprechen.

5.5.2 Diskussion ausgewählter Ergebnisse

a) Zu den Verteilungen der zentralen KFA-Variablen (AERGO, AERGM)

Bei den zentralen KFA-Variablen (AERGO, AERGM) finden sich keine verbotenen Werte. Die Verteilungen fallen unterschiedlich aus. Einen recht normalen Eindruck macht die Verteilung der Ärgermessung in der Situation *ohne* kontrafaktische Alternative (AERGO):



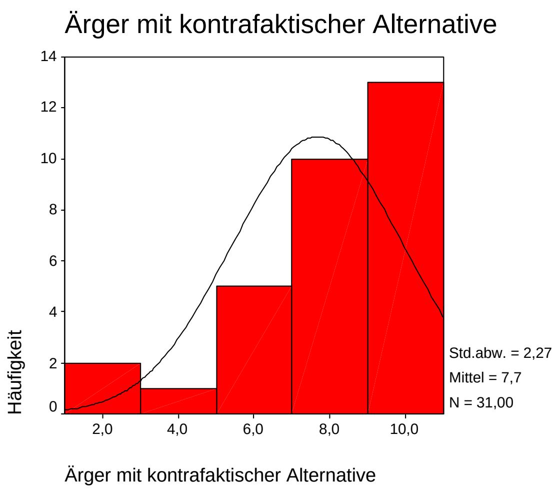
Die Verteilungskennwerte Schiefe (= -0,08) und Kurtosis (= -0,277) sind nach den oben angegebenen Tests nicht signifikant von 0 verschieden:

Statistiken

		Ärger ohne kontrafaktische Alternative	Ärger mit kontrafaktischer Alternative
N	Gültig	31	31
	Fehlend	0	0
Mittelwert		5,68	7,68
Median		6,00	8,00
Modus		5[a]	8
Standardabweichung		2,01	2,27
Varianz		4,03	5,16
Schiefe		-,080	-1,451
Standardfehler der Schiefe		,421	,421
Kurtosis		-,277	2,013
Standardfehler der Kurtosis		,821	,821

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

Die Verteilung der Ärgermessung in der Situation *mit* kontrafaktischer Alternative (AERGM) sieht deutlich anders aus:



Diese Verteilung ist am rechten Rand konzentriert und deutlich verschieden von einer Normalverteilung, was sich auch in signifikanten Ergebnissen der Tests zu Schiefe und Kurtosis widerspiegelt:

$$\frac{|\,\text{Schiefe}\,|}{\text{SF(Schiefe)}} = 3{,}447 > 1{,}96$$

$$\frac{|\text{Kurtosis}|}{\text{SF(Kurtosis)}} = 2{,}451 > 1{,}96$$

Abweichend von meiner generellen Vorliebe für gerichtete Tests habe ich hier *zwei*seitig getestet, weil keine gerichteten Hypothesen vorlagen.

b) Ergebnis der Fehleranalyse

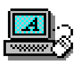
Unsere Fehleranalyse liefert nur einen „Treffer“. In der Häufigkeitstabelle zur Variablen LOT10 entdecken wir den verbotenen Wert 0:

LOT10

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0	1	3,2	3,2	3,2
	1	4	12,9	12,9	16,1
	2	10	32,3	32,3	48,4
	3	9	29,0	29,0	77,4
	4	7	22,6	22,6	100,0
	Gesamt	31	100,0	100,0	

Diese Fehlerquote kann als erfreulich niedrig eingestuft werden.

5.6 Suche nach Daten

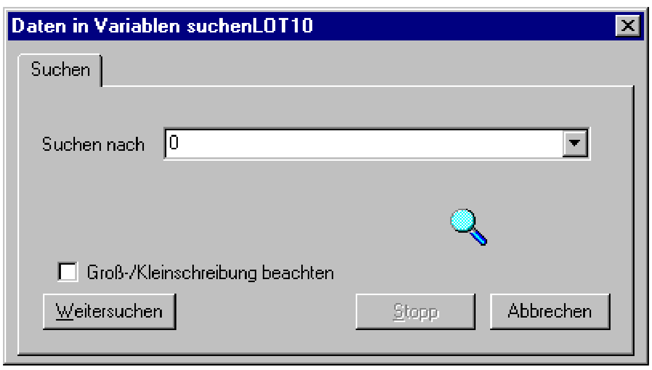


In der Häufigkeitstabelle zur LOT10 haben wir den unzulässigen Wert 0 (mit Häufigkeit 1) entdeckt. Nun möchten wir natürlich sofort wissen, bei welchem Fall dieser Wert auftritt, um geeignete Korrekturen vornehmen zu können. Der betroffene Fall ist sehr leicht zu ermitteln:

- Holen Sie nötigenfalls das Datenfenster in der Vordergrund.
- Markieren Sie in der **Datenansicht** eine beliebige Zelle der Variablen LOT10.
- Klicken Sie auf das Symbol, oder wählen Sie den Menübefehl:

  Bearbeiten > Suchen...

  Dann erscheint die folgende Dialogbox:



- Tragen Sie den zu suchenden Wert ein, und klicken Sie auf den Schalter **Weitersuchen**. Für die Suche nach SYSMIS ist ein Leerzeichen einzutragen.

- Daraufhin markiert SPSS die erste Trefferzelle, und Sie kennen den Fall mit fehlerhaftem LOT10-Wert:
  Es ist zufällig der erste Fall (FNR = 1), dessen ausgefüllter Fragebogen im Manuskript wiedergegeben ist (siehe Seite 21), so dass Sie den korrekten Wert ablesen und im Datenfenster eintragen können. Nach dieser Datenkorrektur sollten Sie die Arbeitsdatei sichern und damit die SPSS-Datendatei **kfar.sav** auf den neuen Stand bringen.

5.7 Arbeiten mit dem Ausgabefenster (Teil II)

Weil es sich beim SPSS Viewer um eine komplexe Anwendung handelt, wird ihre umfangreiche Funktionalität in mehreren Portionen präsentiert.

5.7.1 Nachbearbeitung von Tabellen

Sie werden noch sehr flexible Möglichkeiten zum Umstrukturieren („Pivotieren“) von Tabellen mit dem sogenannten Pivot-Editor kennen lernen (z.B. Zeilen- und Spaltendimension vertauschen). Zunächst beschränken wir uns auf Gestaltungsmöglichkeiten, die das Erscheinungsbild einer Tabelle beeinflussen, ohne ihre Grundstruktur zu verändern. Auch für solche Nachbearbeitungen ist der Pivot-Editor zuständig.
Als Beispiel soll im folgenden die Häufigkeitstabelle zur Fachbereichsvariablen verwendet werden:

Fachbereich an der Universität Trier

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	I	19	61,3	61,3	61,3
	III	2	6,5	6,5	67,7
	IV	6	19,4	19,4	87,1
	VI	4	12,9	12,9	100,0
	Gesamt	31	100,0	100,0	

5.7.1.1 Pivot-Editor starten

Um das Editieren einer Tabelle zu beginnen, können Sie einen Mausdoppelklick darauf setzen oder die Option **Objekt: SPSS Pivot-Tabelle** aus ihrem Kontextmenü wählen. Bei der letztgenannten Methode bietet ein Untermenü die Auswahl zwischen dem **Bearbeiten** innerhalb des Viewers („in-place-editing“) und dem **Öffnen** eines separaten Fensters für das Editieren der Tabelle.

5.7.1.2 Modifikation von Zellinhalten

a) Text editieren

Bei aktivem Pivot-Editor können Sie nach einem Doppelklick auf eine Zelle den enthaltenen Text beliebig ändern. Wir wollen den Titel und die Spaltenbeschriftungen ändern sowie das Wort „Gültig“ am linken Rand der Tabelle löschen:

Fachbereiche im SPSS-Kurs

	n	%	gültige %	kum %
I	19	61,3	61,3	61,3
III	2	6,5	6,5	67,7
IV	6	19,4	19,4	87,1
VI	4	12,9	12,9	100,0
Gesamt	31	100,0	100,0	

Mit der Pivot-Funktion **Gruppierung aufheben** werden wir übrigens später eine Möglichkeit kennen lernen, die überflüssige Zelle mit der Beschriftung „Gültig“ komplett zu entfernen.

b) Zellen zur weiteren Bearbeitung markieren

Mit dem Menübefehl **Bearbeiten > Auswählen** lassen sich Tabellenbestandteile (z.B. Tabellenkorpus, Datenzellen) zur weiteren Bearbeitung markieren.
Außerdem stehen die windows-üblichen Markierungsmethoden per Maus und Tastatur zur Verfügung.

c) Schriftmerkmale

Für eine oder mehrere markierte Zellen kann man nach **Format > Schriftart...** diverse Schriftmerkmale ändern.

d) Zelleneigenschaften

Nach **Format > Zelleneigenschaften** können zahlreiche Attribute der markierten Zellen beeinflusst werden, z.B.:

- Zahlenformate, Anzahl der Dezimalstellen
- Ausrichtung der Zellinhalte
- Randabstände der Zellinhalte
- Schattierung

Mit zentrierten Werten, zwei Dezimalstellen bei den Prozentangaben und rechtsbündig gesetzten Fachbereichsbezeichnungen sieht unsere Beispieltabelle folgendermaßen aus:

Fachbereiche im SPSS-Kurs

	n	%	gültige %	kum %
I	19	61,29	61,29	61,29
III	2	6,45	6,45	67,74
IV	6	19,35	19,35	87,10
VI	4	12,90	12,90	100,00
Gesamt	31	100,00	100,00	

e) Spaltenbreite

Wenn sich der Mauszeiger über dem rechten Rand einer Spalte befindet, ändert er seine Form zu einem doppelseitigen Pfeil. Jetzt können Sie durch Klicken und Ziehen bei gedrückter linker Maustaste die rechte Spaltenbegrenzung verschieben und somit die Spaltenbreite ändern. Der Menübefehl

Ansicht > Gitterlinien

blendet Hilfslinien an der Stelle unsichtbarer Zellenbegrenzungen ein (bzw. aus) und erleichtert damit die Anpassung der Spaltenbreiten.
In unserer Beispieltabelle kann die erste Spalte eine Schlankheitskur vertragen:

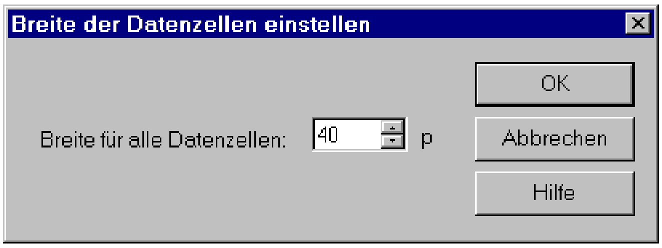
Fachbereiche im SPSS-Kurs

	n	%	gültige %	kum %
I	19	61,29	61,29	61,29
III	2	6,45	6,45	67,74
IV	6	19,35	19,35	87,10
VI	4	12,90	12,90	100,00
Gesamt	31	100,00	100,00	

Über den Menübefehl

Format > Breite der Datenzellen...

lässt sich die Breite sämtlicher Datenzellen einer Tabelle numerisch spezifizieren, z.B.:



Nach missratenen Gestaltungsbemühungen bringt eventuell

Format > Automatisch anpassen

wieder ein akzeptables Ergebnis zu Stande.

5.7.1.3 Tabellenvorlagen

Für eine Pivot-Tabelle kann nach **Format > Tabellenvorlagen...** das Design einer Tabellenvorlage übernommen werden. So sieht unser Beispiel nach Anwendung der Vorlage **Akademisch** aus:

Fachbereiche im SPSS-Kurs

	n	%	gültige %	kum %
I	19	61,29	61,29	61,29
III	2	6,45	6,45	67,74
IV	6	19,35	19,35	87,10
VI	4	12,90	12,90	100,00
Gesamt	31	100,00	100,00	

5.7.2 Weitere Gestaltungsmöglichkeiten im Navigationsbereich

5.7.2.1 Blöcke bzw. Teilausgaben kopieren, verschieben oder löschen

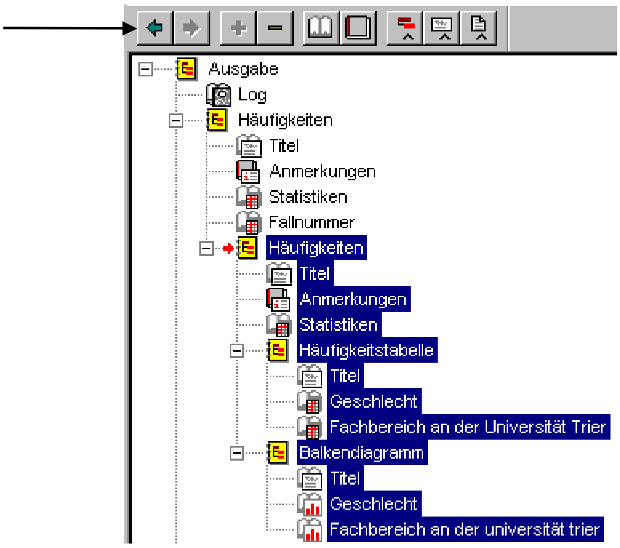
Sie können markierte Blöcke bzw. Teilausgaben ...

- Löschen: mit der **<Entf>**-Taste
- Kopieren bzw. Verschieben:
  - mit der Maus: Ziehen und Fallenlassen, beim *Kopieren* zusätzlich *nach* Beginn der Bewegung die **<Strg>**-Taste drücken
  - via Zwischenablage: **Kopieren** bzw. **Ausschneiden** und **Einfügen** im Menü **Bearbeiten**

Es ist u.a. auch möglich, eine Teilausgabe in einen anderen Ausgabeblock zu kopieren oder zu verschieben.

5.7.2.2 Befördern und Degradieren

Wenn kopierte oder verschobene Ausgabeblöcke versehentlich auf einer unerwünschten Gliederungsebene gelandet sind, können sie mit den Pfeiltasten oberhalb der Navigationszone „befördert“ oder „degradiert“ werden, z.B.:



Die Ausgabeblöcke in einem Viewer-Dokument müssen nicht unbedingt nebeneinander auf der selben Gliederungsebene liegen, sondern können baumartig angeordnet werden. Von dieser Strukturierungsmöglichkeit macht z.B. auch die SPSS-Prozedur zur Häufigkeitsanalyse Gebrauch.

5.7.3 Tabellen im RTF-Format via Zwischenablage in andere Anwendungen übertragen

Wenn Sie SPSS-Tabellen in der Zielanwendung überarbeiten oder mit dem Layout des Zieldokumentes versehen wollen, ist die in Abschnitt 5.3.4 beschrieben Übertragung im Grafikformat nicht geeignet. Befördern Sie in dieser Situation die Tabellen vom SPSS-Viewer aus mit dem Menübefehl **Bearbeiten > Kopieren** bzw. mit der Tastenkombination **<Strg><C>** in die Zwischenablage. In der Zielanwendung landet die Tabelle nach **Bearbeiten > Einfügen** bzw. **<Strg>+<V>** dann im RTF-Format.

Die folgende FB-Häufigkeitstabelle wurde auf die beschriebene Weise in Microsoft Word© übertragen:

Fachbereiche im SPSS-Kurs

		n	%	gültige %	kum %
	I	19	61,29	61,29	61,29
	III	2	6,45	6,45	67,74
	IV	6	19,35	19,35	87,10
	VI	4	12,90	12,90	100,00
	Gesamt	31	100,00	100,00	

Man erhält eine normale Word-Tabelle mit abgeschalteten (hier gestrichelt dargestellten) Gitterlinien, die auf beliebige Weise überarbeitet werden kann, ohne dass dabei Verschlechterungen des Designs zu befürchten sind. Allerdings sind bis zum zufriedenstellenden Endergebnis noch einige Design*verbesserungen* erforderlich. Erste Erfolge bei geringem Arbeitsaufwand können Sie in Word (97 oder 2000) z.B. mit dem Befehl **Tabelle > Tabelle AutoFormat...** erzielen. Nach Zuweisung des Autoformats **Standard 1** und Löschen der überflüssigen ersten Spalte sieht die FB-Häufigkeitstabelle folgendermaßen aus:

Fachbereiche im SPSS-Kurs

	n	*%*	*gültige %*	*kum %*
I	19	61,29	61,29	61,29
III	2	6,45	6,45	67,74
IV	6	19,35	19,35	87,10
VI	4	12,90	12,90	100,00
Gesamt	31	100,00	100,00	

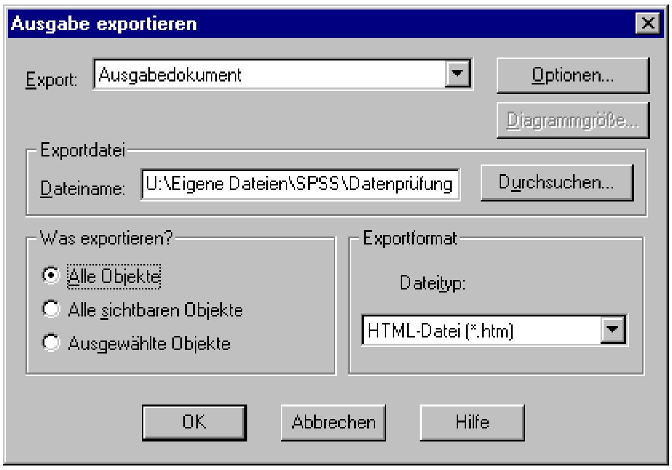
Bei komplizierteren Tabellen führt die Übertragung im RTF-Format allerdings oft zu unbrauchbaren Ergebnissen.

5.7.4 Ausgaben exportieren

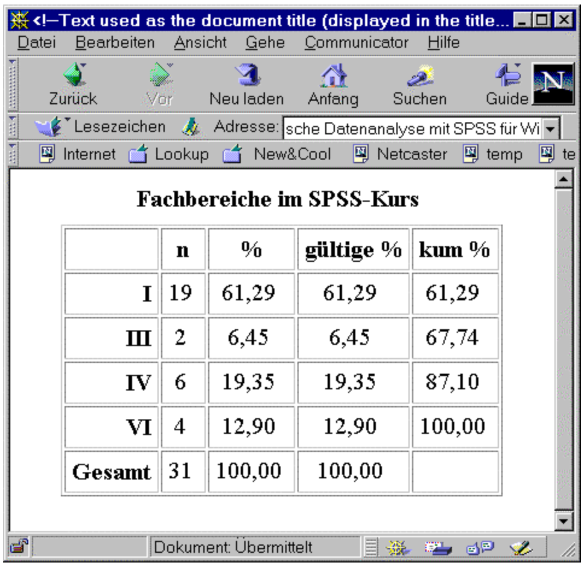
Pivot-Tabellen, Diagramme und Textausgaben können im HTML- oder Textformat exportiert werden. Der Export wird angefordert mit

Datei > Exportieren...

Mit folgender Dialogbox wird z.B. das gesamte Viewer-Dokument im HTML-Format exportiert:



So sieht die oben schon mehrfach betrachtet FB-Häufigkeitstabelle nach dem HTML-Export im Netscape-Browser aus[1]:



[1] Dabei habe ich ein wenig gemogelt, indem ich die überflüssige erste Spalte mit Methoden beseitigt habe, die Sie erst später kennen lernen werden.

Beim Export-Umfang gibt es folgende Alternativen:

- **Ausgabedokument**
- **Ausgabedokument (ohne Diagramme)**
- **Nur Diagramme**
  Dann sind folgende **Dateitypen** zulässig: CGM, JPG, PCT, PNG, EPS, TIF, BMP, WMF

Für jedes zu exportierende Diagramm wird eine eigene Datei erstellt. Beim Exportumfang **Ausgabedokument** können die oben genannten Dateiformate (CGM, JPG etc.) in der **Optionen**-Subdialogbox eingestellt werden. In Abhängigkeit vom gewählten Grafik-Dateityp sind für den Export von Diagrammen weitere Optionen vorhanden, z.B. zur Größe und Farbumsetzung.

5.7.5 Mehrere Ausgabefenster verwenden

Bislang war immer von *dem* Ausgabefenster die Rede. Im Verlauf einer längeren Auswertungsarbeit kann es im Sinne besserer Übersichtlichkeit sinnvoll sein, ein zusätzliches Ausgabefenster anzufordern. Dazu dient der Menübefehl:

Datei > Neu > Ausgabe

Wenn mehrere Ausgabefenster vorhanden sind, muss geregelt werden, in welches Fenster SPSS zukünftige Ausgaben schreiben soll. Daher ist stets ein *Haupt*ausgabefenster festgelegt (gelegentlich auch vornehm als „designiert“ bezeichnet), das durch ein Ausrufezeichen in seiner Statuszeile gekennzeichnet ist. Außerdem ist der Ausrufezeichen-Schalter  in der *Symbol*leiste des Hauptfensters notwendigerweise inaktiv. Dieser Schalter dient nämlich ggf. dazu, ein Ausgabefenster zu designieren.
Um ein bestimmtes Ausgabefenster in den Vordergrund zu holen, können Sie es anklicken oder das **Fenster**-Menü eines beliebigen SPSS-Fensters benutzen.
Jedes Ausgabefenster kann auf windows-übliche Weise geschlossen werden, z.B. indem Sie es in den Vordergrund holen und dann anordnen:

Datei > Schließen

6 Speichern der SPSS-Kommandos zu wichtigen Anweisungsfolgen

6.1 Zur Motivation

Eventuell möchten Sie nach zahlreichen Datenkorrekturen alle Testprozeduren erneut durchführen, um ein beruhigendes Ergebnis „Null Fehler" zu sehen. Leider müssen dazu zahlreiche Dialogboxen erneut ausgefüllt und abgeschickt werden. In diesem Abschnitt werden Sie lernen, wie man eine solche Misere verhindern kann, indem man wichtige Anweisungsfolgen zur späteren Wiederverwendung speichert.

In fast jedem Projekt treten solche wichtigen, potentiell mehrfach benötigten, Anweisungsfolgen auf. Es ist z.B. typisch für empirische Studien, dass vor der eigentlichen Auswertung aus den Rohvariablen mit zahlreichen Datentransformationen neue oder modifizierte „Fertig"-Variablen erstellt werden müssen. In unserem KFA-Projekt muss etwa zur Untersuchung der differentialpsychologischen Hypothese eine Optimismus-Variable aus den 12 LOT-Variablen gebildet werden. Die für ein Projekt erforderlichen Datentransformationen können in SPSS für Windows durch Bearbeiten einer mehr oder weniger langen Serie von Dialogboxen erledigt werden. Auch bei dieser Anweisungsfolge ist eine Konservierung ratsam, denn sie hat folgende Eigenschaften:

- Die einzelnen Anweisungen sind relativ komplex und damit ebenso fehleranfällig wie zeitaufwendig.
- Es ist relativ wahrscheinlich, dass die gesamte Anweisungsfolge wiederholt durchgeführt werden muss. Wenn z.B. bei einer schriftlichen Befragung die Antworten schleppend eintreffen, möchte man vielleicht die bereits vorhandenen Fragebögen vorab auswerten, aber so, dass nicht nach Vervollständigung der Stichprobe alle Arbeiten wiederholt werden müssen.
  In diesem Zusammenhang ist auch zu beachten, dass in SPSS abgeleitete Variablen (z.B. Summen-Variablen) bei einer Änderung der zugrundeliegenden Variablen **nicht** automatisch aktualisiert werden, sondern ihre alten Werte behalten. Folglich müssen nach Datenkorrekturen die betroffenen Berechnungsanweisungen für abgeleiteten Variablen erneut ausgeführt werden.
- Die Anweisungen zur Datentransformation sind „dokumentationspflichtig".
  Damit das Projekt transparent und reproduzierbar bleibt, muss genauestens protokolliert werden, welche Datentransformationen in welcher Reihenfolge durchgeführt wurden. Z.B. muss im KFA-Projekt dokumentiert werden, wie die Optimismus-Variable aus den eingetippten LOT-Variablen berechnet worden ist. Spätestens nach dem Auftreten unsinniger Ergebnisse muss die *tatsächlich* angewendete Berechnungsvorschrift als mögliche Fehlerquelle überprüft werden.

Anweisungssequenzen mit diesen Eigenschaften sollten unbedingt auf irgendeine Weise abgespeichert werden, damit die abgelaufenen Aktionen jederzeit nachvollzogen werden können, und damit nötigenfalls die gesamte Sequenz mit geringem Aufwand wiederholt werden kann.

In SPSS eignen sich zur Konservierung wichtiger Anweisungssequenzen in natürlicher Weise die **Kommandos**, die den einzelnen Dialogboxen zugrunde liegen, und die von SPSS stets im Hintergrund erzeugt und ausgeführt werden, wenn wir eine ausgefüllte Dialogbox mit **OK** abschicken.

In diesem Zusammenhang lohnt sich ein kurzer Blick auf die Architektur des SPSS-Systems, das aus den beiden folgenden Komponenten besteht:

- **Benutzerschnittstelle**
  Wir interagieren mit der Benutzerschnittstelle, die unsere Anweisungen entgegennimmt und die Ergebnisse präsentiert. Wir können der Benutzerschnittstelle unsere Anweisungen in Form von ausgefüllten Dialogboxen oder als Folge von SPSS-Kommandos übergeben.
- **SPSS-Prozessor**
  Die Benutzerschnittstelle gibt unsere Anweisungen in jedem Fall in Form von SPSS-Kommandos an den Prozessor weiter, der im Hintergrund arbeitet. Wir erfahren übrigens in der Statuszeile der SPSS-Fenster, was der Prozessor gerade treibt. Da wir den Prozessor bislang nur minimal belastet haben, hat die Statuszeile meistens angezeigt: **SPSS Prozessor bereit**. Während der Prozessor arbeitet, wird in der Statuszeile protokolliert, mit welchem SPSS-Kommando er gerade beschäftigt ist. Nach dem Abschicken einer Häufigkeitsdialogbox erscheint z.B. **Ausführen: FREQUENCIES**, bei unserem kleinen Datensatz allerdings nur sehr kurz. Wenn wir eine ausgefüllte Häufigkeitsdialogbox mit **OK** quittieren, führt der SPSS-Prozessor also im Hintergrund das korrespondierende FREQUENCIES-Kommando aus.

In fast allen SPSS-Dialogboxen können über die Standardschaltfläche **Einfügen** die zugrunde liegenden SPSS-Kommandos produziert werden. Diese werden dann *nicht* ausgeführt, sondern in ein sogenanntes **Syntaxfenster** übertragen, das weitgehend analog zu einem Texteditor funktioniert. Hier können Sie alle Kommandos zu einer Sequenz ansammeln, nach Bedarf einzeln oder geschlossen ausführen lassen und schließlich in einer Datei abspeichern. Später können Sie die Kommandos aus dieser Datei wieder laden und, eventuell nach manueller Überarbeitung, erneut ausführen lassen. Das genaue Vorgehen wird in Abschnitt 6.2 an einem konkreten Beispiel geübt.
Eine Folge von SPSS-Kommandos kann man (leicht hochstaplerisch) als **SPSS-Programm** bezeichnen. In fast jedem Projekt sollte es mindestens *ein* SPSS-Programm geben, nämlich das in Abschnitt 3 bereits vorgeschlagene Transformationsprogramm, das aus der Rohdatendatei durch diverse Transformationen die Fertigdatendatei des Projektes erstellt. Wir werden für unser KFA-Projekt ein solches Programm in Abschnitt 7 erstellen.
Ob sich bei einer konkreten Anweisungssequenz das Abspeichern als SPSS-Programm lohnt, muss von Fall zu Fall nach obigen Kriterien entschieden werden. Bei kurzen, simplen Sequenzen mit geringer Wiederholungswahrscheinlichkeit ist ein Konservieren sicher unrentabel.
Es soll nicht verschwiegen werden, dass die Ausführung einer Anweisungssequenz mit dem Umweg über ein Syntaxfenster geringfügig mehr SPSS-Kenntnisse erfordert als die direkte Ausführung durch Quittieren der Dialogboxen mit **OK**. Wer sich beim Umgang mit SPSS-Kommandos unsicher fühlt, bei seinem relativ kleinen Projekt eventuell erforderliche Wiederholungen von Dialogbox-Sequenzen nicht scheut und das Risiko inkonsistenter Datenzustände durch große Sorgfalt kontrolliert, der kann auf das Erzeugen und Abspeichern von SPSS-Kommandos verzichten.

Für ambitionierte SPSS-Anwender muss noch klargestellt werden, dass die Erstellung, Überarbeitung und Ausführung von Programmen in einem Syntaxfenster eine eigenständige Methode der SPSS-Benutzung darstellt, über die fast alle Analyse-Funktionen erreichbar sind. Viele SPSS-Leistungen stehen sogar *ausschließlich* über die Syntax zur Verfügung, z.B.:

- Conjoint-Analyse
- Kontrollstrukturen wie z.B. DO REPEAT - Schleifen, mit denen man komplexe Datentransformationen auf effiziente Weise durchführen kann.
- Die MATRIX-Programmiersprache, mit der man eigene Statistikprozeduren programmieren kann.

Der Hersteller SPSS Inc. meint im Hilfesystem zu der Debatte „Dialogbox kontra Programm“:

> „Erfahrene SPSS-Anwender bevorzugen möglicherweise die rationellere Befehlssprache.“

Dazu eine kleine Illustration: Bei der Variablendefinition für unser KFA-Projekt mussten wir die Namen LOT01 bis LOT12 einzeln eintragen. Mit dieser Methode die Variablennamen FPI001 bis FPI287 vereinbaren zu müssen, ist eine schwer erträgliche Vorstellung. Es gibt zwar Schlimmeres, aber es gibt auch Besseres, nämlich den folgenden Dreizeiler, der alle 287 Namen fehlerfrei erzeugt[1]:

```
data list free /fpi001 to fpi287.
begin data.
end data.
```

Im aktuellen Abschnitt 6 werden der Einfachheit halber nur sehr oberflächliche Hinweise zur Kommandosprache gegeben. Diese sollten genügen für Anwender, die nicht frei programmieren, sondern nur gelegentlich ein von SPSS automatisch erzeugtes Kommando modifizieren wollen. Der Anhang enthält für ambitionierte SPSS-Anwender eine ausführlichere Beschreibung der Kommandosprache.
Wie schon erwähnt, sind die Dialogboxen beim Erstellen eines SPSS-Programms sehr nützlich. Mit Hilfe der bislang ignorierten Standardschaltfläche **Einfügen** kann nämlich die zu einer Dialogbox-Bearbeitung äquivalente Kommandofolge in ein Syntaxfenster übertragen werden. Sie müssen sich also nicht zwischen zwei unvereinbaren SPSS-Bediensystemen entscheiden, sondern sollten eine möglichst effiziente Kombination beider Methoden verwenden.

6.2 Dialogunterstützte Erstellung von SPSS-Programmen

Angenommen, Sie haben die Fehleranalyse für Ihre Daten mit Hilfe mehrerer Häufigkeitsanalysen abgeschlossen. Nun erhalten Sie wider Erwarten noch zahlreiche weitere bearbeitete Fragebögen von Untersuchungsteilnehmern, die zunächst nicht mitwirken wollten. Sie freuen sich natürlich über die Stichprobenerweiterung und erfassen sofort die neuen Fälle. Dann allerdings fällt Ihnen ein, dass nun alle Kontrollanalysen wiederholt werden, also alle Häufigkeits-Dialogboxen neu ausgefüllt werden müssen. Solche unangenehmen Situationen kann man durch Konservieren der zugrunde liegenden Kommandos vermeiden.
Das folgende SPSS-Programm führt für unser KFA-Projekt die Häufigkeitsanalysen zur Fehlersuche bei den Variablen FNR, GESCHL und FB durch (siehe Abschnitt 5):

[1] Man kann z.B. folgendermaßen vorgehen, um den gewünschten Effekt zu erreichen:
- Syntaxfenster öffnen mit **Datei > Neu > Syntax**.
- Obige Kommandos eintragen.
- Alle Zeilen markieren und ausführen lassen mit **<Strg><R>**.

```
GET
  FILE='U:\Eigene Dateien\SPSS\KFAR.SAV'.
FREQUENCIES
  VARIABLES=fnr
  /STATISTICS=MINIMUM MAXIMUM MODE
  /ORDER=  ANALYSIS.
FREQUENCIES
  VARIABLES=geschl fb
  /BARCHART  FREQ
  /ORDER=  ANALYSIS.
```

Wir werden dieses Programm gleich „vollautomatisch“ mit drei Mausklicks auf den **Einfügen**-Schalter produzieren und dabei auch seine Bestandteile kurz erläutern. Dazu benötigen wir eine aktive SPSS-Sitzung mit leerem Datenfenster. Starten Sie nötigenfalls SPSS bzw. entleeren Sie das Datenfenster mit:

Datei > Neu > Daten

Rufen Sie die Dialogbox zum Öffnen einer Datendatei auf:

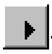
Datei > Öffnen > Daten

Schreiben oder klicken Sie den Namen Ihrer Rohdatendatei in das entsprechende Feld, und betätigen Sie dann den Schalter **Einfügen**. Daraufhin beginnt SPSS *nicht* damit, aus der angegebenen Datendatei eine neue Arbeitsdatei zu erstellen und diese im Datenfenster anzuzeigen, sondern SPSS schreibt das für diese Aktion zuständige GET-Kommando in ein Syntaxfenster mit dem Titel **Syntax1**:



Der Aufbau des GET-Kommandos ist sehr einfach:

- Es beginnt mit dem Kommandonamen GET.
- Im FILE-Subkommando wird die zu öffnende Datei spezifiziert.
- Am Ende muss wie bei jedem SPSS-Kommando ein **Punkt** stehen.

Weil das Datenfenster noch leer ist, stehen die Menübefehle zum Anfordern von Statistik- und Grafikprozeduren noch nicht zur Verfügung. Daher wollen wir jetzt das GET-Kommando ausführen lassen, um die Daten einzulesen. Setzen Sie dazu die Schreibmarke an eine beliebige Position *innerhalb des GET-Kommandos*, und klicken Sie auf das Symbol . Nun erstellt SPSS eine neue Arbeitsdatei mit Ihren Rohdaten und zeigt diese im Datenfenster an.
Spezifizieren Sie jetzt mit Hilfe der zuständigen Dialogbox dieselbe Häufigkeitsanalyse zur FNR-Variablen wie in Abschnitt 5. Verlassen Sie die Dialogbox jedoch nicht mit **OK**, sondern mit **Einfügen**. Daraufhin erscheint am Ende des Syntaxfensters ein FREQUENCIES-Kommando (siehe oben):

- Es beginnt mit dem Kommandonamen FREQUENCIES.
- Im VARIABLES-Subkommando ist angegeben, welche Variable analysiert werden soll.
- Im STATISTICS-Subkommando ist angegeben, welche Verteilungskennwerte berechnet werden sollen.
- Das (im vorliegenden Fall irrelevante) ORDER-Subkommando entscheidet bei der Analyse mehrerer Variablen darüber, ob die Statistiken für jede Variable in einer eigenen Tabelle oder für alle Variablen in einer gemeinsamen Tabelle ausgegeben werden sollen.

Um diese Entscheidung in der **Häufigkeiten**-Dialogbox zu treffen, müssen Sie übrigens die **Format**-Subdialogbox öffnen und im Rahmen **Mehrere Variablen** die passende Option wählen.

- Das FREQUENCIES-Kommando wird wie jedes SPSS-Kommando durch einen Punkt abgeschlossen.

Produzieren Sie als nächstes die Syntax zu der in Abschnitt 5 durchgeführten Häufigkeitsanalyse für die Variablen GESCHL und FB.

Nun sollte Ihr Syntaxfenster den zu Beginn des Abschnitts wiedergegebenen Inhalt haben. Das GET-Kommando ist schon gelaufen, folglich müssen Sie noch die beiden FREQUENCIES-Kommandos ausführen lassen. Weil es sich um *zwei* Kommandos handelt, müssen Sie folgendermaßen vorgehen:

- Markieren Sie zunächst per Maus *die beiden* auszuführenden Kommandos.
- Klicken Sie dann auf das Symbol ▶, oder drücken Sie die Tastenkombination **<Strg><R>**. Daraufhin werden alle Kommandos im Syntaxfenster ausgeführt, die (zumindest teilweise) markiert sind.

Im Ausgabefenster protokolliert SPSS übrigens zu jedem Ausgabeblock in der zunächst zugeklappten Teilausgabe **Anmerkungen** u.a. die zugrunde liegende Syntax und zwar sowohl bei direkter Anforderung einer Leistung per Dialogbox als auch bei programm-orientierter Arbeitsweise, z.B.:

Anmerkungen

Ausgabe erstellt		27-MAR-2001 23:49:35
Kommentare		
Eingabe	Daten	U:\Eigene Dateien\SPSS\kfar.sav
	Filter	<keine>
	Gewichtung	<keine>
	Aufgeteilte Datei	<keine>
	Anzahl der Zeilen in der Arbeitsdatei	31
Behandlung fehlender Werte	Definition von fehlenden Werten	Benutzerdefinierte fehlende Werte werden als fehlend behandelt.
	Verwendete Fälle	Statistik basiert auf allen Fällen mit gültigen Daten.
Syntax		FREQUENCIES VARIABLES=geschl /BARCHART FREQ /ORDER= ANALYSIS .
Ressourcen	Zugelassene Werte	18724
	Verstrichene Zeit	0:00:00,02

Damit sich durch spätere Wiederverwendung der SPSS-Kommandos der Rationalisierungseffekt der programm-orientierten Arbeitsweise einstellen kann, müssen Sie Ihr SPSS-Programm sichern.
Wechseln Sie dazu in das Syntaxfenster, und wählen Sie den Menübefehl:

Datei > Speichen unter...

Verwenden Sie im Dateinamen die vorgeschlagene Erweiterung **.sps**, an der SPSS die Syntaxdateien erkennt.

Wenn Sie später z.B. mit einer erweiterten Stichprobe dieselbe Auswertung nochmals durchführen wollen, dann müssen Sie lediglich das vorhandene Programm mit dem Menübefehl:

Datei > Öffnen > Syntax

laden, komplett markieren (z.B. mit **<Strg><A>**) und erneut ausführen lassen (z.B. mit **<Strg><R>**).

6.3 Arbeiten mit dem Syntax-Fenster

Das Syntaxfenster bietet die Funktionalität eines Texteditors, so dass Sie automatisch erstellte SPSS-Kommandos beliebig modifizieren können. Z.B. werden Sie sicher auch ohne systematische Syntaxkenntnisse in der Lage sein, ein Statistikkommando für andere Variablen ausführen zu lassen. Dazu brauchen Sie lediglich die Variablennamen auszutauschen.
Sie können ein neues Syntaxfenster auch unabhängig von der **Einfügen**-Schaltfläche direkt anfordern mit:

Datei > Neu > Syntax

Wenn *mehrere* Syntaxfenster vorhanden sind, muss geregelt werden, in welches Fenster SPSS die per **Einfügen**-Schalter automatisch erzeugten Kommandos übertragen soll. Dies geschieht genauso wie bei den Ausgabefenstern: Ein Mausklick auf den aktiven Schalter  in seiner Symbolleiste macht ein Syntaxfenster zum designierten. Ein passiver (nicht verwendbarer) Schalter signalisiert ebenso wie ein Ausrufezeichen in der *Status*zeile: Das aktuelle Syntaxfenster *ist* designiert.
Um ein bestimmtes Syntaxfenster in den Vordergrund zu holen, können Sie es anklicken oder das **Fenster**-Menü eines beliebigen SPSS-Fensters benutzen.
Jedes Syntaxfenster kann auf windows-übliche Weise geschlossen werden, z.B. indem Sie es in den Vordergrund holen und dann anordnen:

Datei > Schließen

Wenn Sie längere Zeit mit SPSS arbeiten, wird sich vermutlich Ihr Umgang mit SPSS-Syntax in folgenden Stufen weiterentwickeln:

- Programme automatisch erzeugen lassen und später unverändert wiederverwenden
  Bei dieser Arbeitsweise müssen Sie nur wissen, wie man SPSS-Kommandos per Dialogbox in ein Syntaxfenster hinein befördert, und wie man unerwünschte Kommandos löscht.
  Sie arbeiten dann sozusagen mit einem bequemen Makrorekorder.
- Automatisch erzeugte Programme modifizieren
  Es zeigt sich, dass *konkrete* SPSS-Kommandos meist leicht zu durchschauen und zu modifizieren sind (siehe obiges Beispiel-Programm).
- Freies Programmieren

6.4 Kurzbeschreibung der SPSS-Syntax

In diesem Abschnitt werden einige wenige Merkmale der SPSS-Syntax genannt, damit Sie beim Zusammenstellen bzw. Modifizieren automatisch erzeugter Kommandos möglichst wenig Fehler machen. Eine ausführlichere Beschreibung finden Sie im Anhang.

Die wichtigsten Regeln für SPSS-Kommandos:

- Ein Kommando besteht aus seinem Namen und den zugehörigen Spezifikationen, die sich aus Schlüsselwörtern wie VARIABLES, Variablennamen usw. zusammensetzen:

Kommandoname

```
FREQUENCIES
     VARIABLES=fnr
     /STATISTICS=MINIMUM MAXIMUM MODE.
```

Spezifikationen

- Zwei Elemente der Kommandosprache sind durch mindestens ein Leerzeichen oder durch einen Zeilenwechsel voneinander zu trennen. Manche Zeichen mit festgelegter Bedeutung wie z.B. "=", "/", "(", "+", ">" sind selbstbegrenzend, d.h. vor und nach ihnen sind keine Leerzeichen nötig (aber erlaubt).

- Ein Kommando kann sich über beliebig viele Fortsetzungszeilen erstrecken, dabei dürfen aber *innerhalb* des Kommandos keine Leerzeilen auftreten. Diese signalisieren nämlich normalerweise (wie der Punkt) das Ende des Kommandos.

- Zwischen zwei Kommandos dürfen beliebig viele Leerzeilen stehen, was eine übersichtliche Gestaltung von SPSS-Programmen erlaubt.

- **Jedes Kommando muss in einer neuen Zeile beginnen und mit einem Punkt enden.**

Gut kommentierte Programme sind später leichter zu verstehen. Das gilt auch für ein nach den Empfehlungen in diesem Manuskript fast vollständig automatisch erzeugtes SPSS-Programm.
Sie können zum Kommentieren das SPSS-Kommando COMMENT verwenden, dessen Name auch durch „*“ ersetzt werden darf. Die Syntax ist denkbar einfach:

COMMENT *kommentar*.

oder:

* *kommentar*.

Für dieses Kommando gelten die dieselben Regeln wie für alle anderen SPSS-Kommandos:

- Es darf sich über beliebig viele Fortsetzungszeilen erstrecken, wobei innerhalb des Kommandos keine Leerzeilen erlaubt sind.
- **Beachten Sie, dass auch jedes Kommentar-Kommando mit einem Punkt abgeschlossen werden muss.** Wenn Sie den Punkt am Ende vergessen, dann betrachtet SPSS den folgenden Programmtext bis zum nächsten Punkt als Teil des Kommentars!
- Endet eine Kommentarzeile mit einem Punkt, so betrachtet SPSS das Kommentar-Kommando als abgeschlossen. Wenn Sie einen Punkt als *Satzzeichen* ans Ende einer Kommentarzeile gesetzt haben, dann müssen Sie die nächste Kommentarzeile wieder mit COMMENT oder „*“ einleiten.
  Punkte innerhalb einer Kommentarzeile sind kein Problem.

Beispiel:

```
* Mit diesem Programm wird die Rohdatendatei KFAR.SAV
  auf Erfassungsfehler untersucht.

GET
  FILE='U:\Eigene Dateien\SPSS\KFAR.SAV'.
       .   .   .
       .   .   .
```

7 Datentransformation

7.1 Vorbemerkungen

Die zur Untersuchung unserer differentialpsychologischen Hypothese benötigte Optimismus-Variable existiert noch gar nicht, sondern muss erst aus den 12 LOT-Variablen berechnet werden. Vor dieser Berechnung müssen allerdings die aus messtechnischen Gründen umgepolten (negativ formulierten) LOT-Fragen geeignet rekodiert werden (z.B. Frage 3). Es ist typisch für empirische Studien, dass vor der eigentlichen Auswertung aus den Rohvariablen mit zahlreichen Datentransformationen neue oder modifizierte „Fertig"-Variablen erstellt werden müssen.

In diesem Abschnitt werden Sie häufig benötigte SPSS-Befehle zur Datentransformation kennen lernen. Diese wirken sich auf die Datenmatrix der Arbeitsdatei aus, wo entweder neue Variablen aufgenommen oder vorhandene Variablen verändert werden. Per Voreinstellung werden dabei *alle Fälle gleichermaßen* behandelt.

Man kann die Ausführung einer Datentransformation aber auch von einer Bedingung abhängig machen, so dass i.a. nicht mehr alle Fälle davon betroffen sind. Diese Möglichkeit werden wir z.B. dazu verwenden, unsere MD-Behandlung bei den Motiv-Variablen in Ordnung zu bringen, indem wir genau für *die* Fälle mit "MOTIV1 = MOTIV2 = ... = KEINE = 0" bei allen genannten Variablen die 0 in SYSMIS umkodieren.

Datenmodifikationen sind für numerische und alphanumerische Variablen möglich. Wir wollen uns allerdings auf den erheblich wichtigeren Fall numerischer Variablen beschränken.

7.1.1 Rohdatendatei, Transformationsprogramm und Fertigdatendatei

In Abschnitt 3 wurde vorgeschlagen, zu jedem Projekt ein SPSS-Transformationsprogramm zu erstellen, dessen Aufgabe darin besteht, ausgehend von der Rohdatendatei alle „Fertig"-Variablen zu entwickeln, die im weiteren Verlauf routinemäßig benötigt werden. *Alle* potentiell relevanten Variablen (roh oder fertig) sollen in einer erweiterten Datendatei gesichert werden, die sich für alle Auswertungsarbeiten eignet[1]. Mit Rücksicht auf diese Idee haben wir die bislang existierende Datendatei mit **kfar.sav** ("r" für "roh") bezeichnet. Im Namen der endgültigen Projekt-Fertigdatendatei können wir das "r" dann weglassen.

Wir werden im Verlauf des aktuellen Abschnitts 7 das SPSS-Transformationsprogramm zu unserem KFA-Projekt sukzessive mit Hilfe verschiedener Dialogboxen erstellen (vgl. Abschnitt 6.2). Dabei ist besondere Sorgfalt geboten, weil fehlerhafte Anweisungen im Transformationsprogramm schwerwiegende Konsequenzen für die weitere Arbeit haben können.

Weil das Transformationsprogramm eventuell wiederholt benötigt wird, z.B. nach einer Stichprobenerweiterung oder nach einer Fehlerkorrektur in den Rohdaten, muss es ebenso sorgfältig *gesichert* werden wie die Rohdatendatei. Als Dateinamen wollen wir **kfat.sps** wählen.

Wie in Abschnitt 6.1 ausführlich diskutiert, können Sie alle erforderlichen Transformationen auch durch direkte Ausführung von Dialogboxen (Schalter **OK**) erledigen. Diese Arbeitsweise ist zweifellos für Anfänger leichter zu handhaben als die programm-orientierte Methode, bei der mit Hilfe von Dialogboxen (Schalter **Einfügen**) zunächst mehrere SPSS-Kommandos in ein

---

[1] Unter gewissen, am ehesten in Großprojekten anzutreffenden Umständen kann es sinnvoll sein, die auszuwertenden Daten in mehreren Dateien bereitzuhalten. In der Regel führt das Verteilen der Variablen oder Fälle auf mehrere Dateien früher oder später zu dem Problem, dass sich die in einer Analyse zu vergleichenden Fälle oder Variablen in verschiedenen Dateien befinden. Daher ist unreflektierte Anwendung der allgemeinen Lebensregel „Teile und herrsche!" auf die Dateiorganisation eines Forschungsprojektes nicht zu empfehlen.

Syntaxfenster befördert werden, um sie anschließend ausführen zu lassen. Die direkte Arbeitsweise hat aber folgende Nachteile:

- Beim sukzessiven manuellen Modifizieren der Datendatei kann man bei größeren Projekten leicht den Überblick verlieren. Z.B. weiß irgendwann von einer bestimmten Variablen niemand mehr, in welchen Zwischenschritten sie aus welchen anderen Variablen berechnet worden ist.
- Sind Wiederholungen von Datenmodifikationen erforderlich, müssen diese komplett neu spezifiziert werden. Solche Wiederholungen sind u.a. deshalb häufig erforderlich, weil SPSS abgeleitete Variablen **nicht** automatisch anpasst, wenn sich Werte der Ursprungsvariablen ändern. Nach Korrekturen bei den Rohvariablen müssen Sie also alle Datenmodifikationen wiederholen, in die diese Rohvariablen eingehen.

7.1.2 Hinweise zum Thema Datensicherheit

Ihre Rohdaten können nach der sorgfältigen Datenerfassung und -prüfung zumindest vorerst, z.B. bis zur Entdeckung weiterer Erfassungsfehler, als korrekt gelten. Sichern Sie den erreichten Stand, indem Sie die Rohdaten in mindestens **zwei** Dateien speichern (z.B. auf verschiedenen Disketten). Es ist sinnvoll, für beide Dateien das Schreibschutz-Attribut zu setzen (z.B. mit dem Windows-Explorer).
Vor der geplanten Änderung einer Datei muss das Schreibschutz-Attribut natürlich wieder aufgehoben werden.
Ähnlich sorgfältig sollten Sie nach seiner Fertigstellung das Transformationsprogramm sichern.

Wenn Sie beim Verlassen von SPSS gefragt werden, ob Sie das Daten- oder ein Syntaxfenster sichern wollen, sollten sie sehr sorgfältig prüfen, ob bei dem entsprechenden Objekt während der Sitzung tatsächlich nur geplante Veränderungen stattgefunden haben.



Antworten Sie im Zweifelsfall eher mit **Nein**. Möglicherweise haben Sie durch unbeabsichtigte Tastendrücke Daten gelöscht oder verändert. Diese Fehler sollten dann auf keinen Fall verewigt werden.

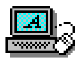
7.1.3 Initialisierung neuer numerischer Variablen

Wenn Sie in einer Datenmodifikations-Anweisung die Erstellung einer *neuen* Variablen anfordern, dann wird die (Fälle × Variablen)-Datenmatrix in der Arbeitsdatei um eine Spalte erweitert. SPSS **initialisiert** dabei zunächst die neue Variable, indem es für alle Fälle den globalen MD-Indikator System-Missing als Wert einträgt. Gelingt anschließend die Ermittlung der neuen Variablenausprägung für einen Fall, so wird der Initialwert entsprechend ersetzt. Anderenfalls bleibt System-Missing stehen, so dass der betroffene Fall bei allen Berechnungen mit der neuen Variablen ausgeschlossen wird.
In diesem Abschnitt sollen Sie nicht belehrt, sondern von einer Sorge befreit werden: Beim Erstellen neuer Variablen hat SPSS alle potentiellen Störungen durch fehlende Werte oder verwandte Probleme im Griff.

7.2 Alte Werte einer Variablen auf neue abbilden (Umkodieren)

Mit dem Befehl **Umkodieren** aus dem Menü **Transformieren** bzw. mit dem äquivalenten RECODE-Kommando können die Werte einer bestehenden Variablen in neue Werte überführt werden. Man kann die Ausgangsvariable verändern oder eine neue Variable mit dem rekodierten Wertevektor erstellen.

7.2.1 Das praktische Vorgehen am Beispiel einer künstlichen Gruppenbildung



Da wir im Abschnitt 7 das KFA-Transformationsprogramm sukzessive aufbauen wollen, öffnen wir zunächst unsere Rohdatendatei **kfar.sav**.

Um das Umkodieren üben zu können, wählen wir ein mäßig sinnvolles Beispiel aus unserer Studie: Wir konstruieren unter dem Namen DEKADE eine vergröberte Variante der Jahrgangsvariablen, bei der alle in den 60'er Jahren geborenen Personen den Wert 1 und alle in den 70'er Jahren geborenen Personen den Wert 2 erhalten sollen. Wie man sich anhand der Häufigkeitstabelle zur Variablen GEBJ überzeugen kann, ist damit für alle Fälle in unserer Stichprobe ein DEKADE-Wert definiert.
Mit Hilfe der neuen Variablen kann man z.B. den Einfluss des Geburtsjahrzehnts auf diverse abhängige Variablen untersuchen, wobei man sich von der Informationsreduktion (im Vergleich zu GEBJ) keinen allzu großen Nutzen versprechen sollte.

Bei der geplanten Rekodierung wird die (Fälle × Variablen)-Datenmatrix um eine neue Variable erweitert, die folgendermaßen aus der vorhandenen Variablen GEBJ entsteht:

GEBJ		DEKADE
69	→	1
70	→	2
69	→	1
67	→	1
.		.
.		.
.		.
72		2
68	→	1
67	→	1
67	→	1

Wählen Sie den Menübefehl:

Transformieren > Umkodieren

Daraufhin erscheint das folgende Untermenü:

In dieselben Variablen...
In andere Variablen...

Da wir eine neue Variable erzeugen wollen, ist die zweite Alternative zu wählen. Machen Sie folgendermaßen weiter:

- Befördern Sie in der nun erscheinenden Dialogbox **Umkodieren in andere Variablen** die Variable GEBJ in das Feld **Eingabevar. -> Ausgabevar**.
- Tragen Sie im Bereich **Ausgabevariable** den gewünschten Namen DEKADE der neu zu erzeugenden Variablen ein.
- Ergänzen Sie als Label „Die größten Hits der 60'er und 70'er Jahre“.
- Drücken Sie dann auf **Ändern**.

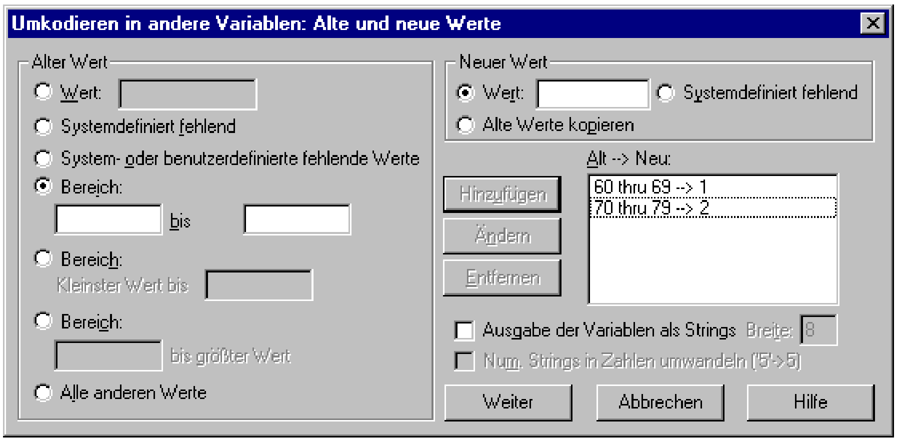
Danach müsste Ihre Dialogbox so aussehen:



Legen Sie nun die Abbildungsregeln fest:

- Aktivieren Sie mit dem Schalter **Alte und neue Werte** die Subdialogbox **Umkodieren in andere Variablen: Alte und neue Werte**.
- Geben Sie unter **Alter Wert** den Bereich von 60 bis 69 an, und benennen Sie als zugehörigen neuen Wert die Eins.
- Beenden Sie die Definition der ersten Abbildungsvorschrift mit **Hinzufügen**.
- Vereinbaren Sie analog die Zuordnungsvorschrift: „70 bis 79 → 2“.

Jetzt müssten Sie dieses Bild sehen:



Damit ist die Rekodierung vollständig spezifiziert. Quittieren Sie die Subdialogbox mit **Weiter**. Da wir das KFA-Transformationsprogramm sukzessive aufbauen wollen, müssen Sie nun in der Dialogbox **Umkodieren in andere Variablen** auf den Schalter **Einfügen** drücken, um die implizit definierten Kommandos zu produzieren. Wir erhalten ein Syntaxfenster mit folgendem Inhalt:

```
RECODE
  gebj
  (60 thru 69=1)  (70 thru 79=2)  INTO  dekade .
VARIABLE LABELS dekade "Die größten Hits der 60'er und 70'er Jahre".
EXECUTE .
```

Neben dem RECODE-Kommando, das die eigentliche Umkodierung bewirkt, hat SPSS noch zwei weitere Kommandos erzeugt:

- VARIABLE LABELS
  Mit diesem Kommando wird das Label für die neue Variable in den Deklarationsteil der Arbeitsdatei eingetragen.
- EXECUTE
  Die Rolle dieses Kommandos wird in Abschnitt 7.3 erläutert.

Offenbar hat SPSS unsere Angaben nur in leicht verständliche, englischsprachige Formulierungen übersetzt, so dass Sie es eigentlich wagen können, die Kommandos bei Bedarf auch in abgeänderter Form zu verwenden.

7.2.2 Technische Details

Obwohl das Umkodieren eine sehr simple Datentransformation ist, sind bei der praktischen Anwendung doch einige technische Details zu beachten:

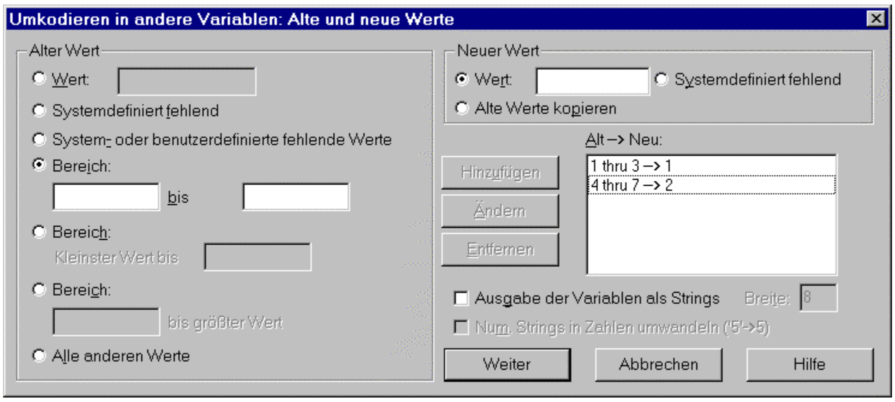
- Sie können bei einem Einsatz der Dialogbox **Umkodieren in andere Variablen** beliebig viele Variablen gleichzeitig umkodieren.
- Bei der Spezifikation der alten Werte, die auf einen neuen Wert abgebildet werden sollen, können Sie angeben:
  - Einen einzelnen **Wert**
  - **Systemdefiniert fehlend** (SYSMIS)
    So ist es also möglich, den automatischen Missing-Data-Indikator auf einen anderen Wert umzusetzen.
  - **System- oder benutzerdefinierte fehlende Werte**
    Alle MD-Indikatoren werden umgesetzt.
  - Den **Bereich** von einem ersten Wert bis zu einem zweiten Wert (jeweils inklusive)
  - Den **Bereich** vom niedrigsten Wert in der Stichprobe bis zu einem bestimmten Wert (inklusive)
    Obwohl hinter SYSMIS faktisch die kleinste im Rechner darstellbare Zahl steckt, wird dieser spezielle Wert dem nach unten unbegrenzten Bereich *nicht* zugerechnet.
  - Den **Bereich** von einem bestimmten Wert bis zum höchsten Wert in der Stichprobe (inklusive)
  - **Alle anderen Werte**
    Damit sind alle in keiner anderen Ersetzungsvorschrift genannten Werte angesprochen (inklusive MD-Indikatoren, auch System-Missing). **Alle anderen Werte** kann nur in *einer* Ersetzungsvorschrift angegeben werden. Diese wird von SPSS in der Liste aller Ersetzungsvorschriften automatisch an die letzte Stelle gesetzt und damit bei der Kommando-Ausführung zuletzt abgearbeitet.
- Als neuen Wert, auf den die alten Werte einer Ersetzungsvorschrift abgebildet werden sollen, können Sie angeben:
  - Einen **Wert**
  - **Systemdefiniert fehlend** (SYSMIS)
    Dann werden alle zugehörigen alten Werte auf SYSMIS umgesetzt.

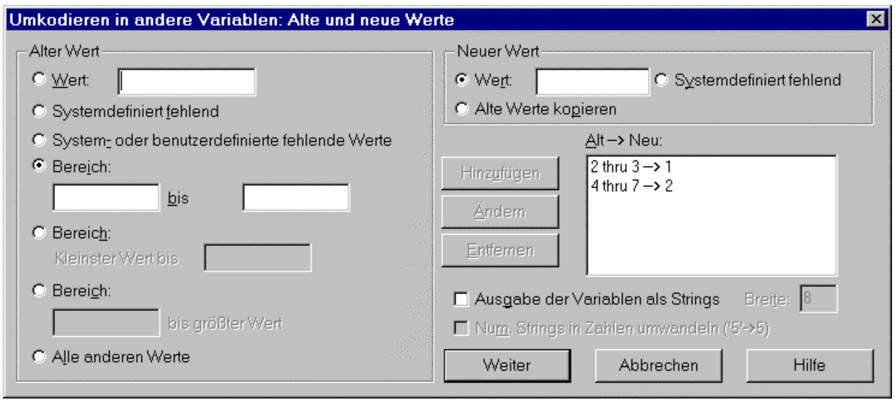
- **Alte Werte kopieren**
  Diese Möglichkeit steht nur beim Umkodieren in *andere* Variablen zur Verfügung und bewirkt dann für die zugehörigen alten Werte eine unveränderte Übernahme. Dies ist besonders nützlich, wenn die alten Werte mit **Alle anderen Werte** spezifiziert worden sind.

- Sie können beliebig viele Ersetzungsvorschriften festlegen. SPSS bringt diese automatisch in eine sinnvolle Ordnung.
- Wenn beim Umkodieren in andere Variablen eine *neue* Variable entsteht, so wird diese zunächst initialisiert, d.h. für alle Fälle wird in der neuen Spalte der Arbeitsdatei der Wert System-Missing eingetragen (vgl. Abschnitt 7.1.3). Durch die *erste zutreffende* Übersetzungsregel wird bei einem Fall der Initialisierungswert durch den zugehörigen neuen Wert überschrieben. Wird der alte Wert eines Falles in keiner Übersetzungsregel angesprochen, dann bleibt bei der neuen Variablen der Initialisierungswert System-Missing stehen! Dies würde in obigem Beispiel etwa einem 1980 geborenen Untersuchungsteilnehmer passieren.
- **Benutzerdefinierte MD-Indikatoren werden wie gültige Werte behandelt!**
  Hat z.B. eine Variable den Wert 9 als benutzerdefinierten MD-Indikator und wird die 9 rekodiert zur 8, dann **bleibt** die 9 ein MD-Indikator der Variablen, insbesondere ist die 8 **nicht** automatisch ebenfalls MD-Indikator. Eventuell muss also nach der Rekodierung die Variablenbeschreibung angepasst werden.
- Jeder Fall wird **nur einmal** umkodiert, und zwar gemäß der **ersten zutreffenden** Ersetzungsregel (bei Anordnung von oben nach unten).

7.2.3 Übungen

1) In den beiden folgenden Dialogboxen, die wir allerdings in unserem Projekt nicht wirklich ausführen wollen, wird jeweils eine Umkodierung der Fachbereichs-Variablen (FB) in eine andere (neue) Variable spezifiziert. Hätten die beiden Dialogboxen den selben Effekt?





2) Bei unserem LOT-Fragebogen wurden die Fragen 3, 4, 8, und 12 aus messtechnischen Gründen umgepolt. Indem eine „optimistische“ Antwort abwechselnd durch Zustimmung oder Ablehnung zum Ausdruck kommt, wird vermieden, dass systematische Ja- oder Nein-Sager einen extremen Optimismuswert erhalten. Bevor wir einen Mittelwert aus den LOT-Fragen als Optimismus-Schätzwert errechnen können, müssen die negativ gepolten Variablen folgendermaßen umkodiert werden:

$$5 \rightarrow 1$$
$$4 \rightarrow 2$$
$$2 \rightarrow 4$$
$$1 \rightarrow 5$$

Arbeiten Sie mit der Umkodieren-Dialogbox, aber quittieren Sie Ihre Eintragungen nicht mit **OK**, sondern mit **Einfügen**, damit das zugehörige RECODE-Kommando in das Syntaxfenster eingetragen wird, in dem wir gerade unser Transformationsprogramm aufbauen.
Machen Sie sich klar, warum die Abbildungsvorschrift „3 → 3“ beim Umkodieren **In dieselben Variablen** überflüssig ist, beim Umkodieren in andere (neue) Variablen aber unbedingt erforderlich wäre.

7.3 Zur Rolle des EXECUTE-Kommandos

Wenn Sie eine **Umkodieren**-Dialogbox mit **OK** quittieren, dann führt SPSS per Voreinstellung die angeforderte Rekodierung sofort in der Arbeitsdatei aus. Obwohl dieses Verhalten sehr sinnvoll und natürlich erscheint, gibt es eine erwägenswerte Alternative. Zum Rekodieren muss SPSS nämlich die Arbeitsdatei vollständig durchlaufen, was bei einer großen Stichprobe durchaus einige Zeit in Anspruch nehmen kann. Bei einer nächsten und übernächsten Transformationsanweisung (z.B. Rekodierung oder Neuberechnung) ist jeweils ein weiterer Durchlauf fällig. Dabei könnte SPSS zeitsparend *alle* Transformationen in einer *einzigen* Datenpassage erledigen. Diese könnte so lange aufgeschoben werden, bis durch die Anforderung einer Statistikprozedur das Durchackern der Daten unvermeidlich wird. Genau in dem zuletzt beschriebenen, ökonomischen Sinn funktionieren seit jeher die SPSS-Transformationskommandos: Sie werden erst bei der nächsten Prozedur ausgeführt. Allerdings kann dieses zeitoptimierte Verhalten SPSS-Neulinge verwirren. Daher setzt SPSS für Windows hinter jedes per Dialogbox implizit (bei Quittieren mit **OK**) oder explizit (bei Quittieren mit **Einfügen**) produzierte Transformationskommando ein EXECUTE-Kommando, welches die *sofortige Ausführung* aller noch offenen Transformationen erzwingt. Wenn wir z.B. eine **Umkodieren**-Dialogbox mit **OK** quittieren, verarbeitet der SPSS-Prozessor im Hintergrund ein RECODE- *und* ein EXECUTE-Kommando. Das erste bewirkt nur eine Arbeitsvorbereitung, das zweite erzwingt die Ausführung der vorbereiteten Arbeit. Quittieren wir dieselbe Dialogbox mit **Einfügen**, erscheinen die beiden Kommandos im Syntaxfenster.

Obwohl bei der in diesem Manuskript vorgestellten Arbeitsweise die von SPSS produzierten EXECUTE-Kommandos in der Regel überflüssig sind, empfehle ich Ihnen **nicht**, diese Kommandos aus den automatisch produzierten Programmen zu entfernen. Bei jeder manuellen Bearbeitung können sich Fehler einschleichen, und der realisierbare Zeitgewinn ist aufgrund der heute verfügbaren Rechenleistung bei den meisten Projekten irrelevant.

Beim Arbeiten mit dem Syntaxfenster kann es leicht zu dem folgenden, recht frustrierenden Erlebnis kommen: Sie lassen ein wohlansehnliches Transformationskommando ausführen und im Datenfenster passiert *Nichts*. Die Ursache ist dann meist: Sie haben nach dem Transformationskommando noch kein Prozedur- bzw. EXECUTE-Kommando ausführen lassen. SPSS zeigt dann in der Statuszeile an, dass **Offene Transformationen** zur Bearbeitung anstehen. Sie

können deren Ausführung ganz einfach erzwingen, indem Sie im Syntaxfenster ein EXECUTE-Kommando abschicken oder folgenden Menübefehl wählen:

Transformieren > Offene Transformationen ausführen

Es soll nicht verschwiegen werden, dass hier für SPSS-Neulinge Schwierigkeiten auftauchen, die bei rein dialogbox-orientierter Arbeitsweise nicht entstehen können.

Man kann übrigens nach

Bearbeiten > Optionen > Daten

im Rahmen **Optionen für Transformieren und Zusammenfügen** mit der Option **Werte vor Verwendung berechnen** die voreingestellte EXECUTE-Inflation abstellen. Dann zeigt SPSS das oben beschriebene zeitoptimierte Verhalten, führt also z.B. nach dem Quittieren einer **Umkodieren**-Dialogbox mit **OK** das zugrunde liegende RECODE-Kommando zunächst noch nicht aus, sondern reiht es in die Warteschlange der offenen Transformationen ein. Diese werden vom SPSS-Prozessor erst dann ausgeführt, wenn er ein Prozedur- oder ein EXECUTE-Kommando erhält.

Für angehende SPSS-Profis möchte ich noch erwähnen, dass EXECUTE-Kommandos *innerhalb eines Blocks von Transformationsanweisungen* durchaus bedeutsam sein können. In dem folgenden (manuell erstellten) Beispiel wird mit Hilfe des Transformationskommandos SELECT IF jeder zweite Fall aus der Arbeitsdatei entfernt:

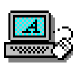
```
compute nr = $casenum.
execute.
select if (mod(nr,2) = 1).
execute.
```

Lässt man das erste ECECUTE weg, entfernt das Programm *alle* Fälle mit Ausnahme des ersten.

7.4 Berechnung von Variablen nach mathematischen Formeln

In der Dialogbox **Variable Berechnen** bzw. im äquivalenten COMPUTE-Kommando wird ein numerischer Ausdruck (z.B.: "GROESSE - 100") definiert und einer Ergebnisvariablen zugewiesen. Dabei kann man eine *neue* Variable erzeugen oder eine vorhandene verändern.

7.4.1 Beispiel



Sie sollen später anhand unserer Stichprobe untersuchen, ob die Trierer Studierenden im Mittel wenigstens das folgende Idealgewicht auf die Waage bringen (Nullhypothese)

$$\textit{Gewicht (in kg)} \overset{!}{=} \textit{Größe(in cm) - 100}$$

oder ob sie relativ zu dieser Formel zu leicht sind (Alternativhypothese). Zur Prüfung dieser Frage mit einem t-Test für gepaarte Stichproben muss die Arbeitsdatei um eine neue Variable, z.B. IDGEW genannt, erweitert werden, deren Werte nach obiger Formel aus der Körpergröße zu berechnen sind. Anschließend enthält die (Fälle × Variablen)-Datenmatrix in der Arbeitsdatei u.a. die beiden folgenden Variablen:

GROESSE	IDGEW
163	63
158	58
174	74
182	82
.	.
.	.
.	.
176	76
176	76
170	70
169	69

Starten Sie zum Definieren der neuen Variablen die Dialogbox **Variable berechnen** mit:

Transformieren > Berechnen...

Tragen Sie zunächst im Feld **Zielvariable** den Namen für die neu in die Arbeitsdatei aufzunehmende Variable ein (IDGEW), und schreiben Sie dann in das Feld **Numerischer Ausdruck** die Definitionsvorschrift (GROESSE - 100), wobei einige Schreibhilfen zur Verfügung stehen:

- Der Variablenname kann aus einer Liste per Transportschalter oder Doppelklick übernommen werden.
- Mit Hilfe der Tastatur auf dem Bildschirm können Sie das Minuszeichen und die Zahl 100 auch per Maus eingeben.

Anschließend sollte Ihre Dialogbox so aussehen:



Die Dialogbox bietet über unsere momentanen Bedürfnisse hinausgehend auch die in SPSS verfügbaren Funktionen (siehe unten) in einer rollbaren Liste zum Transport in das Feld **Numerischer Ausdruck** an, so dass wir bei der Verwendung von Funktionen nicht nachschlagen müssen und weniger Tippfehler machen.

Rufen Sie nun mit dem gleichnamigen Schalter die Subdialogbox **Typ und Label** auf, und tragen Sie dort für IDGEW das Etikett „Idealgewicht nach der Formel: Größe - 100" ein:



Quittieren Sie die Subdialogbox mit **Weiter** und die Hauptdialogbox mit **Einfügen**. Daraufhin erhalten Sie im Syntaxfenster die folgenden Kommandos:

```
COMPUTE idgew = groesse - 100 .
VARIABLE LABELS idgew 'Idealgewicht nach der Formel: Größe - 100' .
EXECUTE .
```

Das der Dialogbox **Variable berechnen** entsprechende COMPUTE-Kommando ist offenbar mit den Erfahrungen aus der Schulmathematik leicht zu durchschauen.

7.4.2 Technische Details

7.4.2.1 Numerischer Ausdruck

Im Bereich **Numerischer Ausdruck** der Dialogbox **Variable berechnen** sind wir trotz der SPSS-Scheibhilfen im wesentlichen wieder in das „Syntaxzeitalter“ zurückgeworfen: Auf der weißen Fläche ist ein sprachlicher Ausdruck nach gewissen Syntaxregeln zu formulieren. Zum Glück sind uns aber numerische Ausdrücke aus der Schule wohlbekannt[1].
Konkret darf ein numerischer Ausdruck im Sinne von SPSS folgende Bestandteile enthalten:

- Bereits definierte Variablen
- Zahlen
- arithmetische Operatoren:
    - Addition (+)
    - Subtraktion (-)
    - Multiplikation (*)
    - Division (/)
    - Potenzfunktion (**)
- Klammern
- Funktionen

7.4.2.1.1 Numerische Funktionen

In numerischen Ausdrücken können Sie zahlreiche Funktionen verwenden, die numerische Variablen oder Zahlen als Argumente (in den folgenden Syntaxdarstellungen vertreten durch den Platzhalter *arg*) verarbeiten.[2] Diese Funktionen lassen sich in mehrere Gruppen einteilen, aus denen jeweils einige wichtige Vertreter genannt werden sollen:

---

1 Zwar gibt es gewisse Unterschiede zwischen mathematischen *Gleichungen* (z.B. y = a + b · x) und EDV-sprachlichen *Zuweisungen* (z.B. compute x = x + 2.), doch sind die Regeln für die numerischen Ausdrücke auf den *rechten* Seiten weitgehend identisch.

2 SPSS kennt auch zahlreiche Funktionen für String- und Datums-Variablen, die aber aus Zeitgründen in diesem Kurs nicht behandelt werden. Informieren Sie sich bei Bedarf im Hilfesystem, z.B. über eine Suche nach dem Stichwort „Funktionen“.

- **Arithmetische Funktionen**, z.B.:

  - ABS(*arg*) Absoluter Wert
  - EXP(*arg*) Exponentialfunktion
  - LG10(*arg*) Dekadischer Logarithmus
  - LN(*arg*) Natürlicher Logarithmus
  - MOD(*arg1*,*arg2*) Rest aus der Division von arg1 durch arg2, z.B.: mod(1.3,1) = 0.3
  - RND(*arg*) Auf eine ganze Zahl gerundeter Wert
  - SQRT(*arg*) Quadratwurzel

  Beispiel:
  ```
  compute logi = exp(3+1.2*x)/(1+exp(3+1.2*x)).
  ```
  Hier wird eine spezielle logistische Funktion der Variablen X definiert.

- **Statistische Funktionen**, z.B.:

  - MEAN[.*n*](*arg1*,*arg2*[, ...]) Arithmetisches Mittel
    Voreinstellung für *n*: 1
  - MAX[.*n*](*arg1*,*arg2*[, ...]) Maximum
    Voreinstellung für *n*: 1
  - MIN[.*n*](*arg1*,*arg2*[, ...]) Minimum
    Voreinstellung für *n*: 1
  - SD[.*n*](*arg1*,*arg2*[, ...]) Standardabweichung
    Voreinstellung für *n*: 2
  - SUM[.*n*](*arg1*,*arg2*[, ...]) Summe
    Voreinstellung für *n*: 1

  Regeln:
  - Die eckigen Klammern schließen optionale Angaben ein.
  - Der Funktionsparameter *n* hat folgende Bedeutung: Wenn bei einem Fall mindestens *n* valide Argumente vorliegen, wird der Funktionswert berechnet. Ansonsten wird dem Fall der Wert SYSMIS zugewiesen. Wenn Sie mit der Voreinstellung für *n* nicht einverstanden sind, können Sie einen alternativen Wert festlegen.
  - Mit „[, ...]“ wird zum Ausdruck gebracht, dass die Liste der Argumente optional beliebig verlängert werden darf.
  - Sie können eine Serie von Variablen, *die in der Arbeitsdatei hintereinander stehen*, bequem auf folgende Weise in einer Argumentenliste angeben:
    erste TO *letzte*
    Es kommt nicht auf die alphanumerische Ordnung der Variablennamen an, sondern auf die tatsächliche Reihenfolge der Variablen in der Arbeitsdatei.

  Beispiel:
  ```
  compute mfrei = mean.45(sport to angeln).
  ```
  Wenn für einen Fall bei den Variablen SPORT bis ANGELN, die in der Arbeitsdatei hintereinander stehen, mindestens 45 valide Argumente vorliegen, wird deren Mittelwert der neuen Variablen zugewiesen, ansonsten wird der MD-Indikator System-Missing zugewiesen.

Beachten Sie den wesentlichen Unterschied zwischen den gerade beschriebenen statistischen *Funktionen* und den Statistik-*Prozeduren*, mit denen wir z.B. oben die Verteilungsanalysen durchgeführt haben:

- Wenn wir in der Dialogbox **Häufigkeiten** (erreichbar über **Analysieren > Deskriptive Statistiken > Häufigkeiten**) z.B. den Mittelwert der Variablen GEWICHT anfordern, werden die (validen) Gewichtsangaben aller Fälle in der Stichprobe gemittelt. Es werden also die Ausprägungen *einer Variablen*

über *alle Fälle* gemittelt. SPSS arbeitet sich *senkrecht* durch eine komplette Variable bzw. Spalte der Arbeitsdatei. Es resultiert ein einziger Stichprobenkennwert, welcher im Ausgabefenster erscheint.

- Mit der statistischen Funktion MEAN können wir für *jede einzelne Person* z.B. den Mittelwert über *mehrere LOT-Variablen* berechnen lassen. SPSS geht *waagerecht* vor, wobei das selbe Verfahren *auf jeden Fall, d.h. auf jede Zeile* der Datenmatrix angewendet wird. Die statistische Funktion MEAN erzeugt (oder modifiziert) eine Variable, d.h. eine komplette Spalte im Datenfenster, in die für jeden Fall sein eigenes Berechnungsergebnis eingetragen wird.

- **Funktionen für fehlende Werte**, z.B.:

- NMISS(*arg1*[, ...])	Anzahl fehlender Werte bei den angegebenen Variablen
- VALUE(*arg*)	Es wird der Wert der Variablen *arg* geliefert, wobei *benutzerdefinierte* MD-Deklarationen ignoriert werden.

Regeln:
- Mit „[, ...]“ wird zum Ausdruck gebracht, dass die Liste der zu untersuchenden Variablen optional beliebig verlängert werden darf.
- Mit dem Schlüsselwort TO können bequem Serien von Variablen angegeben werden (siehe 1. Beispiel und obige Erläuterungen zu den statistischen Funktionen).

Beispiele:
- `compute nmfrei = nmiss(sport to angeln).`
  Der numerische Ausdruck liefert die Anzahl der fehlenden Werte (System- oder Benutzer-Missing) bei den Variablen SPORT bis ANGELN, die in der Arbeitsdatei hintereinander stehen.
- `compute vala = value(a).`
  Diese Funktion liefert auch dann den Wert der Variablen A, wenn es sich um einen benutzerdefinierten MD-Indikator handelt.

- **Pseudozufallszahlengeneratoren**, z.B.:

- NORMAL(*arg*)	Die Funktion liefert normalverteilte Zufallszahlen mit Mittelwert 0 und Standardabweichung *arg*.
- UNIFORM(*arg*)	Die Funktion liefert gleichverteilte Zufallszahlen im Intervall von 0 bis *arg*.

Beispiel:

```
COMPUTE av = NORMAL(1) .
EXECUTE .
T-TEST
  GROUPS=geschl(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=av
  /CRITERIA=CIN(.95) .
```

Die Kommandos aus diesem Beispiel wurden durch Dialogboxen erzeugt. Im COMPUTE-Kommando wird die normalverteilte Zufallsvariable AV definiert. Wir können davon ausgehen, dass unsere Geschlechts-Teilstichproben bis auf Zufallsschwankungen den selben Mittelwert bei AV haben. Damit können wir ausprobieren, wie sich der t-Test zum Vergleich der Mittelwerte zweier unabhängiger Stichproben bei Gültigkeit der Nullhypothese verhält. Die Dialogbox zu diesem t-Test erhält man mit **Analysieren > Mittelwerte vergleichen > t-Test bei unabhängigen Stichproben**.

Wenn Ihnen die Erläuterungen zu diesem Beispiel „spanisch“ vorkommen, hilft Ihnen vielleicht der Abschnitt 8.1 weiter, wo einige Grundprinzipien der Inferenzstatistik erläutert werden.

Hinweis: Bei NORMAL und UNIFORM wird ein Pseudozufallszahlengenerator verwendet, der per Voreinstellung mit dem festen Wert 2000000 startet und damit stets die selben Zahlen liefert. Ein alternativer Startwert, der andere Zufallszahlen liefert, kann gewählt werden:

- mit dem Menübefehl:
  Transformieren > Startwert für Zufallszahlen...
- oder mit dem SPSS-Kommando:
  SET SEED=*n*.

7.4.2.1.2 Regeln für die Bildung numerischer Ausdrücke

Auch bei Verwendung der Dialogbox **Variable berechnen** müssen wir die numerischen Ausdrücke im wesentlichen selbst formulieren. Dabei sind folgende Regeln zu beachten:

- Die **Auswertungsreihenfolge** hängt von der Priorität der Operatoren ab. Es gilt folgende Rangordnung:

Priorität 1:	Funktionen
Priorität 2:	Potenzfunktion (**)
Priorität 3:	Multiplikation (*), Division (/) und Vorzeichen-Minus (z.B.: "-b")
Priorität 4:	Addition (+), Subtraktion (-)

  Bei gleicher Priorität erfolgt die Auswertung von links nach rechts.
  Eine andere Auswertungsreihenfolge kann durch Klammern erzwungen werden: Klammerausdrücke werden zuerst ausgewertet. Bei geschachtelten Klammern erfolgt die Auswertung von innen nach außen.

- Bei Funktionen mit mehreren Argumenten müssen die einzelnen Argumente **durch jeweils genau ein Komma** (optional ergänzt durch Leerzeichen) getrennt werden.
  Beispiel: `compute mabc = mean(a, b,c).`

- Obwohl SPSS *im Daten- und im Ausgabefenster* das ländertypische Dezimaltrennzeichen benutzt, bei uns also das Komma, müssen in numerischen Ausdrücken gebrochene Zahlen generell mit Dezimal**punkt** geschrieben werden.
  Richtig: `2.75`
  Falsch: `2,75`
  Dies gilt sowohl für das Feld **Numerischer Ausdruck** der Dialogbox **Variable berechnen** als auch für das COMPUTE-Kommando in einem Syntaxfenster.
  Es kann also durchaus passieren, dass Sie ein und dieselbe gebrochene Zahl im Datenfenster (als Wert eines Falles für eine bestimmte Variable) mit Dezimal*komma* und in der Dialogbox **Variable berechnen** (z.B. als Konstante in einer Berechnungsanweisung) mit Dezimal*punkt* schreiben müssen.

- In der Regel sind numerische Ausdrücke als Argumente von Funktionen zugelassen.
  Beispiel: `compute albmax = max(a, ln(b)).`
  Das zweite Argument der Funktion MAX ist der numerische Ausdruck `ln(b)`.

7.4.2.2 Sonstige Hinweise

SYSMIS als Ergebnis eines numerischen Ausdrucks

Durch eine Berechnungsanweisung wird der Wert des numerischen Ausdrucks auch dann der Zielvariablen zugewiesen, wenn dieser Wert gleich SYSMIS ist (z.B. bei fehlenden Argumenten). Dieses Vorgehen ist kompatibel mit dem in Abschnitt 7.1.3 beschriebenen Initialisierungsprinzip für *neue* numerische Variablen. Ist die Zielvariable bereits *vorhanden*, bleibt bei missglückter Berechnung des numerischen Ausdrucks keinesfalls der alte Wert stehen, sondern es wird sinnvollerweise SYSMIS zugewiesen.

Rechnen mit fehlenden Werten

Wenn bei einem Fall eine Variable aus dem numerischen Ausdruck keinen validen Wert hat, dann erhält die Ergebnisvariable den Wert SYSMIS. Ausnahmen sind die folgenden SPSS-eigenen Regeln für das „Rechnen" mit fehlenden Werten:

- 0 * unbekannt = 0
  Diese Regel ist schlau, denn:
  $$0 \cdot x = 0 \quad \forall x \in \Re$$
- 0 / unbekannt = 0
  Diese Regel ist kritisierbar, denn:
  $$\frac{0}{x} = \begin{cases} 0 & \text{für } x \neq 0 \\ \text{undefiniert} & \text{für } x = 0 \end{cases}$$
- unbekannt^0 = 1
  Der unbekannte Wert könnte bei einem Fall zufällig gleich Null sein. 0^0 ist aber nicht definiert. SPSS ermittelt hingegen das Ergebnis 1.
- $0^{\text{unbekannt}}$ = 0
  Der unbekannte Wert könnte bei einem Fall zufällig gleich Null sein. 0^0 ist aber nicht definiert. SPSS ermittelt diesmal das Ergebnis 0.
- MOD (0,unbekannt) = 0
  Diese Regel ist kritisierbar, denn:
  $$\text{mod}(0, x) = \begin{cases} 0 & \text{für } x \neq 0 \\ \text{undefiniert} & \text{für } x = 0 \end{cases}$$

Die mathematischen Einwände gegen manche Regeln sollten Ihnen den Schlaf nicht rauben. Jedoch würde ich ein Verkehrsflugzeug nicht betreten, wenn ich wüsste, dass in seinem Steuerprogramm solche „Zahlentricks" verwendet werden.

7.4.3 Übungen

1) Welche Werte haben die folgenden numerischen Ausdrücke?

```
(3 + 4) / 2
 3 + 4  / 2
 3**2 / 2  + 4
(3**2 / 2) + 4
```

2) Erstellen Sie im KFA-Projekt die Variablen, auf die sich unsere zentralen Projekt-Hypothesen beziehen:

- Berechnen Sie die Variable LOT als arithmetisches Mittel der (nötigenfalls rekodierten!) LOT-Variablen 1, 3, 4, 5, 8, 9, 11 und 12. Die restlichen Fragen dienen nicht zur Messung von Optimismus, sondern sollen als sogenannte „Füller" verhindern, dass der

Zweck des Fragebogens deutlich wird. Dies könnte das Antwortverhalten verzerren. Tolerieren Sie bei der Berechnung des Mittelwertes bis zu zwei fehlende Werte.

- Berechnen Sie die Variable AERGAM als arithmetisches Mittel der beiden Ärgervariablen und die Variable AERGZ als Ärgerzuwachs auf Grund der kontrafaktischen Alternative. AERGAM benötigen wir zum Testen der differentialpsychologischen Hypothese. Beim geplanten Test der allgemeinpsychologischen Hypothese wird letztlich mit einem Einstichproben-t-Test geprüft, ob der Mittelwert der Variablen AERGZ signifikant größer als 0 ist. Man kann den Test zwar bequem mit der SPSS-Prozedur zum t-Test für gepaarte Stichproben durchführen, ohne die Variable AERGZ explizit berechnen zu müssen, doch bietet diese Prozedur keine Möglichkeit, die Verteilungsvoraussetzung des Tests zu prüfen. Daher berechnen wir AERGZ explizit und prüfen die Verteilungsvoraussetzung mit der Prozedur zur explorativen Datenanalyse (siehe unten).

Rufen Sie jeweils mit dem Menübefehl:

Transformieren > Berechnen...

die zuständige Dialogbox auf, aber quittieren Sie Ihre Eintragungen nicht mit **OK**, sondern mit **Einfügen**, damit die zugehörigen COMPUTE-Kommandos als Bestandteile des Transformationsprogramms in das designierte Syntaxfenster eingetragen werden.

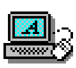
7.5 Bedingte Datentransformation

Häufig ist es erforderlich, eine Datenmodifikation auf diejenigen Fälle zu beschränken, die eine bestimmte Bedingung erfüllen. Wir benötigen z.B. im KFA-Projekt eine solche Möglichkeit, um bei den Motivations- und Methodenvariablen das bisher vertagte Problem der fehlenden Werte adäquat behandeln zu können (siehe Abschnitt 1).
Manchmal ist es angebracht, für mehrere disjunkte Teilmengen der Gesamtstichprobe jeweils spezifische Modifikationen durchzuführen (Fallunterscheidung). Z.B. könnte man im Rahmen einer Untersuchung zum Essverhalten bei der Berechnung der neuen Variablen Idealgewicht aus der bereits vorhandenen Variablen Körpergröße bei Frauen und Männern unterschiedliche Formeln anwenden.

Die Transformations-Dialogboxen enthalten in der Regel die Schaltfläche **Falls**, deren Subdialogbox die Angabe einer Bedingung erlaubt, unter der eine Transformation ausgeführt werden soll. Sie können z.B. eine bedingte Umkodierung (vgl. Abschnitt 7.2), Berechnung (vgl. Abschnitt 7.4) oder Werteauszählung (vgl. Abschnitt 7.6) vornehmen.

Falls unter ein und derselben Bedingung gleich *mehrere* Transformationen vorgenommen werden sollen, muss diese Bedingung für alle benötigten Transformationen, also in allen benötigten Dialogboxen, wiederholt werden, was lästig werden kann. Ebenso sind Fallunterscheidungen mit den Transformations-Dialogboxen leider nur recht umständlich zu realisieren.
Für solche Aufgaben bietet die SPSS-Kommandosprache bessere Lösungen. Diese lassen sich jedoch nicht komplett mit Dialogboxen generieren, so dass sie in diesem Kurs aus Zeitgründen nicht behandelt werden.

7.5.1 Beispiel

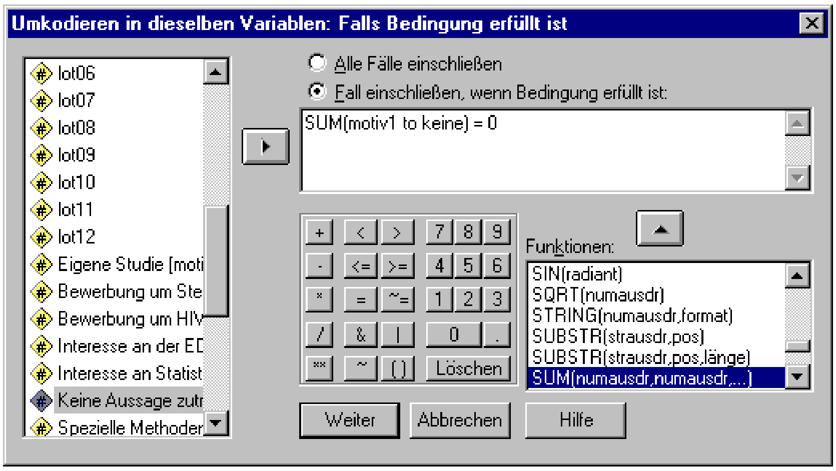


In diesem Abschnitt soll endlich das MD-Problem bei den Motivationsvariablen gelöst werden. Wir haben bei den Variablen MOTIV1 bis MOTIV5 und KEINE systematisch die angekreuzten Kästchen mit 1 und die leeren Kästchen mit 0 kodiert, um während der Erfassung möglichst wenige zeitraubende und fehleranfällige Entscheidungen treffen zu müssen. Ein Fall mit Nullen bei MOTIV1 bis MOTIV5 *und* KEINE ist aber nicht auswertbar, weil er weder Motive angegeben, noch die Abwesenheit aller Motive bestätigt hat. Daher müssen für genau

diese Fälle die Nullen bei den Variablen MOTIV1 bis MOTIV5 und KEINE in System-Missing umkodiert werden. Gehen Sie folgendermaßen vor:

- Wählen Sie den Menübefehl:

  Transformieren > Umkodieren > in dieselben Variablen...
- Transportieren Sie die Variablennamen MOTIV1 bis MOTIV5 und KEINE in die Teilnehmerliste der **Umkodieren**-Dialogbox.
- Legen Sie in der Subdialogbox **Alte und neue Werte** die benötigte Abbildungsvorschrift fest.
- Öffnen Sie die **Falls**-Subdialogbox, markieren Sie die Option **Fall einschließen, wenn Bedingung erfüllt ist**, und tragen Sie in das darunter liegende Textfeld eine geeignete Bedingung ein, z.B.:



Aufgrund unserer Datenüberprüfung können wir uns darauf verlassen, dass bei den Variablen MOTIV1 bis MOTIV5 und KEINE ausschließlich die Werte 0 und 1 vorliegen. Daher ist die Summe dieser Variablen genau dann gleich 0, wenn jede einzelne Variable gleich 0 ist.
Die obige Eintragung im Bedingungsfeld kann „semiautomatisch“ z.B. folgendermaßen erzeugt werden:
  - Markieren Sie in der Funktionenliste **SUM(NumAusdr,NumAusdr,...)** und klicken Sie auf den zugehörigen Transportschalter.
  - Transportieren Sie aus der Variablenliste MOTIV1 in das Bedingungsfeld.
  - Schreiben Sie des Rest der Einfachheit halber per Hand.
- Machen Sie **Weiter** und quittieren Sie die Hauptdialogbox mit **Einfügen**.

Daraufhin wird Ihr Transformationsprogramm um die folgende Sequenz erweitert:

```
DO IF (SUM(motiv1 to keine) = 0) .
RECODE
  motiv1 motiv2 motiv3 motiv4 motiv5 keine  (0=SYSMIS)  .
END IF .
EXECUTE .
```

Wenn Sie diese Kommandos ausführen lassen, gleichgültig ob direkt per **OK** in der **Umkodieren**-Dialogbox oder indirekt via Syntaxfenster, passiert bei jedem einzelnen Fall in der Stichprobe folgendes:

- SPSS prüft die Bedingung, die wir auch als **logischen Ausdruck** bezeichnen wollen.
- Ist bei einem Fall die Bedingung erfüllt, dann wird umkodiert, anderenfalls passiert Nichts.

Weil die Variablen MOTIV1 bis MOTIV5 und KEINE vor der Rekodierung garantiert nur Nullen oder Einsen als Werte aufweisen, hat unser logischer Ausdruck übrigens die Eigenschaft, in jedem Fall entweder wahr oder falsch zu sein. Das erscheint nach dem aussagenlogischen Axiom vom ausgeschlossenen Dritten als selbstverständlich, ist es aber in der empirischen Forschung z.B. wegen des allgegenwärtigen Problems fehlender Werte keineswegs. Für die Fälle in unserer Stichprobe kann z.B. der logischer Ausdruck „GESCHL = 1“ folgende Wahrheitswerte annehmen:

- wahr ⇔ Der GESCHL-Wert ist gleich 1.
- falsch ⇔ Der GESCHL-Wert ist eine von 1 verschiedene Zahl.
- unbestimmt ⇔ Der GESCHL-Wert fehlt, ist also gleich SYSMIS.

Komplexere logische Ausdrücke (z.B. „LN(ML)/ANZ > 1“) können auch wegen undefinierter Funktionswerte unbestimmt sein (z.B. ML ≤ 0 oder ANZ = 0).

Wenn Sie eine bedingte Transformationsanweisung verwenden, sollten Sie beachten, wie SPSS auf bestimmte und unbestimmte logische Ausdruck reagiert:

- Ist der logische Ausdruck **wahr**, dann wird die Transformation ausgeführt.
  Im Fall einer bedingten Berechnung wird der Ergebnisvariablen also der Wert des numerischen Ausdrucks zugewiesen. Die Zuweisung erfolgt auch dann, wenn der numerische Ausdruck den Wert System-Missing hat.
- Ist der logische Ausdruck **falsch oder unbestimmt**, so passiert **nichts**, d.h.:
  - Eine bereits vorhandene Ergebnisvariable behält für den betreffenden Fall ihren bisherigen Wert.
  - Bei einer neu definierten Variablen behält der betroffene Fall den Initialisierungswert System-Missing.

7.5.2 Bedingungen formulieren

Der in obigem Beispiel aufgetretene logische Ausdruck war recht einfach aufgebaut, weil er nur aus einem einzigen Vergleich bestand. Obwohl Ihnen auch komplexe logische Ausdrücke (z.B. aus der Schule) wohlvertraut sein dürften, soll der Begriff zur Klärung einiger Spezialprobleme etwas genauer beschrieben werden. Zunächst wird der einfachere Begriff „Vergleich“ erklärt:

7.5.2.1 Vergleich

Ein Vergleich ist eine besonders einfach aufgebaute Aussage, bestehend aus zwei numerischen Ausdrücken und einem Vergleichsoperator:

numerischer_ausdruck vergleichs-operator numerischer_ausdruck

Die bekannten Vergleichsoperatoren können in SPSS alternativ durch „ASCII-Varianten“ der mathematischen Symbole oder durch Schlüsselwörter dargestellt werden:

Symbol	Schlüsselwort	Bedeutung
=	EQ	gleich
<>	NE	ungleich
<	LT	kleiner als
<=	LE	kleiner oder gleich
>	GT	größer als
>=	GE	größer oder gleich

Beispiel: `beruf > 4`

7.5.2.2 Logischer Ausdruck

Aus dem einfachen Begriff „Vergleich“ wird nun durch eine rekursive Definition der komplexere Begriff „logischer Ausdruck“ konstruiert:

i) Jeder Vergleich ist ein logischer Ausdruck.

ii) Durch Anwendung des logischen Operators **NOT** auf einen logischen Ausdruck oder durch Anwendung der logischen Operatoren **AND** bzw. **OR** auf zwei logische Ausdrücke entsteht ein neuer logischer Ausdruck:

NOT *logischer_ausdruck*

logischer_ausdruck_1 AND *logischer_ausdruck_2*

logischer_ausdruck_1 OR *logischer_ausdruck_2*

Den Wahrheitswert eines zusammengesetzten logischen Ausdrucks erhält man aus den Wahrheitswerten der Argumente nach den Regeln für logische Operatoren, die in den sogenannten Wahrheitstafeln festgelegt sind (siehe unten).

So lassen sich sukzessiv beliebig komplexe logische Ausdrücke aufbauen, die für einen konkreten Fall immer die Wahrheitswerte *wahr*, *falsch* oder *unbestimmt* haben können.

Beispiel: `(lie1 = 0) and (lie2 = 0)`

Das Problem unbestimmter Werte in logischen Ausdrücken löst SPSS analog zu den Regeln für das Rechnen mit fehlenden Werten in numerischen Ausdrücken (siehe Abschnitt 7.4.2.2). Die folgenden Wahrheitstafeln sind gegenüber der klassischen Aussagenlogik um den Wahrheitswert „unbestimmt“ erweitert (*la1* und *la2* seien logische Ausdrücke):

la1	NOT *la1*
wahr	falsch
falsch	wahr
unbestimmt	unbestimmt

la1	*la2*	*la1* AND *la2*	*la1* OR *la2*
wahr	wahr	wahr	wahr
wahr	falsch	falsch	wahr
wahr	unbestimmt	unbestimmt	wahr
falsch	wahr	falsch	wahr
falsch	falsch	falsch	falsch
falsch	unbestimmt	falsch	unbestimmt
unbestimmt	wahr	unbestimmt	wahr
unbestimmt	falsch	falsch	unbestimmt
unbestimmt	unbestimmt	unbestimmt	unbestimmt

7.5.2.3 Regeln für die Auswertung logischer Ausdrücke

Bei der Auswertung von logischen Ausdrücken in SPSS gelten folgende Regeln:

- Die Abarbeitungsreihenfolge hängt von der Priorität der Operatoren ab. Es gilt folgende Rangordnung:

  Priorität 1: Funktionen
  Priorität 2: Potenzfunktion (**)
  Priorität 3: Multiplikation (*), Division (/),
  Vorzeichen-Minus (z.B. -a)
  Priorität 4: Addition (+), Subtraktion (-)
  Priorität 5: Vergleichsoperatoren
  Priorität 6: NOT
  Priorität 7: AND
  Priorität 8: OR
- Bei gleicher Priorität: Abarbeitung von links nach rechts.
- Eine andere Auswertungsreihenfolge kann durch Klammern erzwungen werden.

Beispiel: Das obige Beispiel für einen logischen Ausdruck kann unter Ausnutzung der voreingestellten Abarbeitungsreihenfolge auch einfacher geschrieben werden:

```
lie1 = 0 and lie2 = 0
```

7.5.3 Übung

Bei den Variablen METH1 bis METH5 haben wir zur Vereinfachung der Erfassung im Kodierplan festgelegt, dass „unbenutzte“ Variablen einfach leer bleiben sollen. Nun wollen wir aber bei Fällen mit regulärem Antwortmuster die System-Missing - Werte durch Nullen ersetzen. Die Null soll z.B. bei der Variablen METH2 bedeuten:

„Die Option, einen zweiten Methodenwunsch zu äußern, wurde nicht genutzt.“

Die folgende Tabelle, die wir in Abschnitt 1 vereinbart haben, legt im einzelnen fest, was unter den möglichen Bedingungskonstellationen geschehen soll:

		Mindestens eine speziell interessierende Methode angegeben?	
		Ja	Nein
SMG	1	METH1 ... METH5: SYMIS → 0 Bem.: Korrektes Antworteverhalten. Variablen zu nicht benutzten Optionen (gem. Kodierplan bisher auf SYSMIS) werden auf 0 umgesetzt.	SMG: 1 → SYMIS Bem.: Irreguläres Antwortever-halten. METH1 bis METH5 behal-ten SYMIS. SMG wird ebenfalls auf SYMIS gesetzt.
	0	SMG: 0 → 1 METH1 ... METH5: SYMIS → 0 Bem.: „Leicht“ irreguläres Antworte-verhalten. Wir sind großzügig und passen SMG an.	METH1 ... METH5: SYMIS → 0 Bem.: Korrektes Antworteverhal-ten. Die Variablen zu allen Optio-nen (gem. Kodierplan bisher auf SYMIS) werden auf 0 umgesetzt.
	SYSMIS	SMG: SYSMIS → 1 METH1 ... METH5: SYMIS → 0 Bem.: „Leicht“ irreguläres Antworte-verhalten. Wir sind großzügig und passen SMG an.	Bem.: Irreguläres Antwortever-halten. Alle Variablen behalten den Wert SYMIS.

In den beiden obersten Zeilen jeder Zelle sind die erforderlichen Korrekturen bei SMG bzw. METH1 bis METH5 angegeben.
Erweitern Sie bitte Ihr Programm **kfat.sps** um passende Transformationsanweisungen.

7.6 Häufigkeit bestimmter Werte bei einem Fall ermitteln

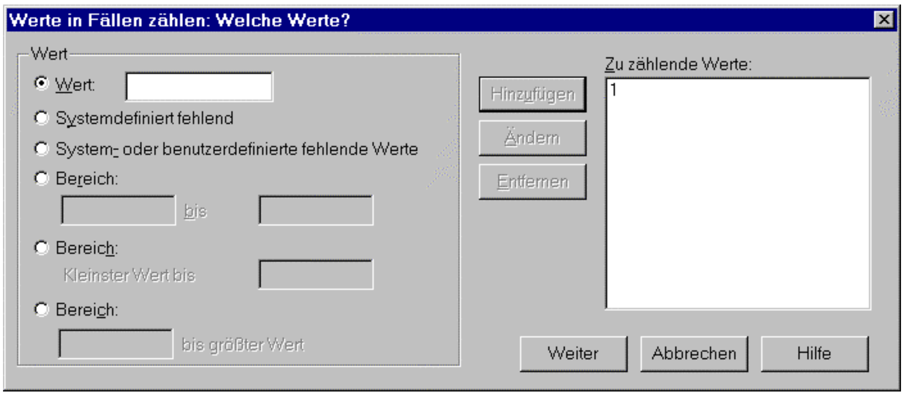
Mit dem Befehl **Zählen** aus dem Menü **Transformieren** bzw. mit dem zugrunde liegenden COUNT-Kommando kann man eine Variable berechnen lassen, die für jeden Fall festhält, wie oft bestimmte Werte in einer Liste von *k* Variablen vorkommen. Das minimale Ergebnis ist 0 (keine Variable hat einen der kritischen Werte), und das maximale Ergebnis ist *k* (jede Variable hat einen kritischen Wert).
Wir wollen eine neue Variable namens POLYMOT berechnen lassen, die für jede Person festhält, wie viele Motive zur Kursteilnahme sie im Fragebogenteil 4a angegeben hat. Aktivieren Sie die Dialogbox **Häufigkeiten von Werten in Fällen zählen** mit

Transformieren > Zählen...

Vergeben Sie für die Zielvariable den Namen POLYMOT und das Label „Anzahl der Motive für die Kursteilnahme“, und transportieren Sie die Variablen MOTIV1 bis MOTIV5 in die Teilnehmerliste. Danach müsste Ihre Dialogbox so aussehen:



Wechseln Sie jetzt mit dem Schalter **Werte definieren** in die Subdialogbox **Werte in Fällen zählen: Welche Werte?**, tragen Sie dort den kritischen Wert Eins ein, und klicken Sie auf **Hinzufügen**:



Die in dieser Subdialogbox angebotenen sonstigen Möglichkeiten zur Festlegung der Trefferwerte kennen wir übrigens schon aus der Subdialogbox **Umkodieren: Alte und neue Werte** (siehe Abschnitt 7.2).

Da SPSS eine Folge von mehreren markierten Kommandos stets in der natürlichen Reihenfolge abarbeitet, wird beim späteren Ablauf unseres Transformationsprogramms die MD-Problematik bei den Variablen MOTIV1 bis MOTIV5 bereits gelöst sein, wenn die **Zählen**-Anweisung an die Reihe kommt. Bei Personen, die den Fragebogenteil 4a nicht korrekt bearbeitet haben, wird also gelten MOTIV1 = MOTIV2 = ... = MOTIV5 = System-Missing.
Wir müssen noch die folgende wichtige Eigenschaft der **Zählen**-Anweisung beachten: Ihre Ergebnisvariable hat *stets* einen validen Wert größer oder gleich 0. Wenn ein Fall z.B. bei allen kritischen Variablen den - nicht zu zählenden - Wert System-Missing hat, resultiert das valide Ergebnis 0! Daher müssen wir, z.B. durch die Bedingung „nmiss(motiv1 to motiv5) = 0" (vgl. Abschnitt 7.4.2.1.1), dafür sorgen, dass ein Fall bei POLYMOT den Wert System-Missing erhält, falls er den Fragebogenteil 4a nicht korrekt bearbeitet hat. In dieser Situation wissen wir nämlich *nichts* von den Motiven der Person, und wir dürfen ihr keine Motivationslosigkeit (POLYMOT = 0) unterstellen.

Klicken Sie daher in der Dialogbox **Häufigkeiten von Werten in Fällen zählen** auf den **Falls**-Schalter, und tragen Sie die eben vorgeschlagenen Bedingung ein. Wenn Sie dann **Weiter** machen und die Hauptdialogbox mit **Einfügen** quittieren, erhalten Sie im Syntaxfenster die folgenden Kommandos:

```
DO IF (nmiss(motiv1 to motiv5) = 0) .
COUNT
  polymot = motiv1 motiv2 motiv3 motiv4 motiv5  (1)   .
VARIABLE LABELS polymot 'Anzahl der Motive für die Kursteilnahme' .
END IF .
EXECUTE .
```

Was hier *zählt*, ist offenbar das COUNT-Kommando. Es enthält im wesentlichen eine Liste der kritischen Variablen, gefolgt von einer eingeklammerten Liste der zu zählenden Werte. Das VARIABLE LABELS - Kommando hat SPSS aufgrund unserer Eintragung im **Label**-Feld erstellt.

7.7 Erstellung der Fertigdatendatei mit dem Transformationsprogramm

Aufgrund der KFA-bezogenen Übungsaufgaben in den Abschnitten 7.2 (Erstellung von DEKADE durch Rekodierung von GEBJ, Umkodieren der negativ formulierten LOT-Fragen), 7.4 (Berechnung von IDGEW, LOT, AERGAM und AERGZ), 7.5 (MD-Behandlung für die Motiv- und für die Methoden-Variablen) und 7.6 (Auszählen der Kursmotive) sollten jetzt alle vorläufig im KFA-Projekt benötigten Transformationskommandos in einem Syntaxfenster stehen.

7.7.1 Transformationsprogramm vervollständigen

Um daraus ein komfortables SPSS-Programm zu machen, das die Rohdatendatei **kfar.sav** selbständig einließt, dann die so entstandene Arbeitsdatei transformiert und schließlich als Fertigdatendatei **kfa.sav** auf die Festplatte sichert, müssen wir an den Anfang des Syntaxfensters noch ein GET-Kommando zum Öffnen von **kfar.sav** und ans Ende noch ein SAVE-Kommando zum Sichern in **kfa.sav** setzen. Wie Sie das GET-Kommando produzieren können, haben Sie schon in Abschnitt 6.2 erfahren. Wenn Sie das Kommando jetzt erzeugen lassen, erscheint es am Ende des Syntaxfensters, und Sie müssen es an den Anfang verschieben. Um das SAVE-Kommando zu generieren, wechseln wir ins Datenfenster und aktivieren mit **Datei > Speichern unter...** die zugehörige Dialogbox. Dann tragen wir den gewünschten Dateinamen **kfa.sav** ein und erzeugen mit Einfügen das benötigte SAVE-Kommando.

Zwei Hinweise zur Ausgabedatei eines Transformationsprogramms:

- Verwenden Sie niemals die selbe Datei als Quelle und Ziel des Transformationsprogramms. Schreiben Sie also keinesfalls mit Ihrem Transformationsprogramm in die Rohvariablendatei. Wenn Sie der Empfehlung in Abschnitt 7.1.2 folgend für die Rohvariablendatei das Schreibschutzattribut gesetzt haben, kann dieses Desaster auch nicht versehentlich passieren.
- Bei der Ausführung des Transformationsprogramms darf für seine Ausgabedatei, also für die Fertigvariablendatei, das Schreibschutzattribut natürlich nicht gesetzt sein.

Schließlich sollte Ihr Syntaxfenster ungefähr so aussehen:

```
GET
  FILE='U:\Eigene Dateien\SPSS\KFAR.SAV'.
EXECUTE .

* DEKADE.
RECODE
  gebj
  (60 thru 69=1)  (70 thru 79=2)  INTO  dekade .
VARIABLE LABELS dekade "Die größten Hits der 60'er und 70'er Jahre".
EXECUTE .

* LOT-Fragen umkodieren.
RECODE
  lot03 lot04 lot08 lot12  (5=1)  (4=2)  (2=4)  (1=5)  .
EXECUTE .

* IDGEW.
COMPUTE idgew = groesse - 100 .
VARIABLE LABELS idgew 'Idealgewicht nach der Formel: Größe - 100' .
EXECUTE .

* LOT berechnen.
COMPUTE lot = MEAN.6(lot01,lot03,lot04,lot05,lot08,lot09,lot11,lot12) .
VARIABLE LABELS lot 'LOT-Optimismus' .
EXECUTE .
```

```
* AERGAM berechnen.
COMPUTE aergam = (aergo + aergm)/2 .
VARIABLE LABELS aergam 'Mittel der Ärger-Variablen' .
EXECUTE .

* AERGZ berechnen.
COMPUTE aergz =  aergm - aergo .
VARIABLE LABELS aergz 'Ärger-Zuwachs durch die KFA' .
EXECUTE .

* MD-Behandlung für die Motiv-Variablen.
DO IF (SUM(motiv1 to keine) = 0) .
RECODE
  motiv1 motiv2 motiv3 motiv4 motiv5 keine  (0=SYSMIS)  .
END IF .
EXECUTE .

* MD-Behandlung für die Methoden-Variablen, Zelle (1,1) der Tabelle.
DO IF (smg=1 and nmiss(meth1 to meth5) < 5) .
RECODE
  meth1 meth2 meth3 meth4 meth5  (SYSMIS=0)  .
END IF .
EXECUTE .

* MD-Behandlung für die Methoden-Variablen, Zelle (1,2) der Tabelle.
DO IF (smg=1 and nmiss(meth1 to meth5) = 5) .
RECODE
  smg  (1=SYSMIS)  .
END IF .
EXECUTE .

* MD-Behandlung für die Methoden-Variablen, Zelle (2,1) der Tabelle.
DO IF ((smg = 0)  and  (nmiss(meth1 to meth5) < 5)) .
RECODE
  smg  (0=1)  .
END IF .
EXECUTE .
DO IF ((smg = 0)  and  (nmiss(meth1 to meth5) < 5)) .
RECODE
  meth1 meth2 meth3 meth4 meth5  (SYSMIS=0)  .
END IF .
EXECUTE .

* MD-Behandlung für die Methoden-Variablen, Zelle (2,2) der Tabelle.
DO IF (smg=0 and nmiss(meth1 to meth5)=5) .
RECODE
  meth1 meth2 meth3 meth4 meth5  (SYSMIS=0)  .
END IF .
EXECUTE .

* MD-Behandlung für die Methoden-Variablen, Zelle (3,1) der Tabelle.
DO IF ((nmiss(smg) = 1)  and  (nmiss(meth1 to meth5) < 5)) .
RECODE
  smg  (SYSMIS=1)  .
END IF .
EXECUTE .
DO IF ((nmiss(smg) = 1)  and  (nmiss(meth1 to meth5) < 5)) .
RECODE
  meth1 meth2 meth3 meth4 meth5  (SYSMIS=0)  .
END IF .
EXECUTE .

* POLYMOT berechnen.
DO IF (nmiss(motiv1 to motiv5) = 0) .
COUNT
  polymot = motiv1 motiv2 motiv3 motiv4 motiv5  (1)  .
VARIABLE LABELS polymot 'Anzahl der Motive für die Kursteilnahme' .
END IF .
EXECUTE .

SAVE OUTFILE='U:\Eigene Dateien\SPSS\KFA.SAV'
   /COMPRESSED.
```

Hierzu sind einige Anmerkungen erforderlich:

- Zwischen manchen Kommandos sind der Übersichtlichkeit halber Leerzeilen eingefügt worden. Man darf aber auf keinen Fall *innerhalb* eines Kommandos (d.h. zwischen dem Kommandonamen und dem abschließenden Punkt) eine Leerzeile einfügen (vgl. Abschnitt 6.4).
- Die mit „*“ eingeleiteten Zeilen beinhalten *Kommentare*, die nachträglich eingefügt wurden, um die spätere Orientierung im Programm zu erleichtern.
  Wichtig: Ein Kommentar hat ebenfalls Kommandostatus und muss daher unbedingt mit einem Punkt abgeschlossen werden. Anderenfalls wird das nächstfolgende Kommando dem Kommentar zugerechnet und folglich nie ausgeführt.
- Das GET-Kommando am Anfang des Programms *überschreibt die aktuelle Arbeitsdatei ohne Nachfrage*! Wenn Sie im Datenfenster manuelle Korrekturen vornehmen, diese nicht sichern, sondern anschließend ein GET-Kommando (via Syntaxfenster) ausführen lassen, dann sind die manuellen Korrekturen verloren.
- Das SAVE-Kommando überschreibt eine eventuell vorhandene Datei **kfa.sav** ohne Nachfrage, was jedoch bei der in diesem Manuskript vorgeschlagenen Arbeitsweise (vgl. Abschnitt 7.1.1) unproblematisch ist.

Eventuell legen Sie Wert darauf, auch bei den neu berechneten Variablen die Attribute Spaltenbreite und Dezimalstellen, die sich auf die Anzeige von Werten auswirken, optimal einzustellen. Dies *manuell* zu tun (vgl. Abschnitt 3.2.2) ist wenig attraktiv, weil unser Transformationsprogramm ja mit einiger Wahrscheinlichkeit mehrfach ausgeführt werden muss. Die bessere Alternative besteht darin, unser Programm um ein FORMATS-Kommando zu erweitern, das die Attribute automatisch setzt:

```
formats dekade (f1.0) idgew (f2.0) aergam (f3.1)
        aergz (f3.0) polymot (f1.0).
```

Im Ausdruck „(f*b*.*d*)“ legt man mit *b* die Gesamtbreite der Wertausgabe (Attribut Spaltenformat) und mit *d* die Anzahl der Dezimalstellen fest. Eine Möglichkeit, die Breite der Datenfensterspalte (Attribut Spalten) per Kommando festzulegen, ist mir leider nicht bekannt.
Offenbar ist das FORMATS-Kommando erst sinnvoll, nachdem alle neu berechneten Variablen bekannt sind. Fügen Sie es daher am Ende des Transformationsprogramm ein, unmittelbar vor dem SAVE-Kommando.

Damit ist das Transformationsprogramm zum KFA-Projekt fertig. Falls noch nicht geschehen, müssen Sie es unbedingt sichern, z.B. in das Verzeichnis **U:\Eigene Dateien\SPSS** unter dem oben vorgeschlagenen Dateinamen **kfat.sps**.

7.7.2 Transformationsprogramm ausführen

Lassen Sie das Transformationsprogramm ausführen. Statt alle Zeilen zu markieren und dann die Ausführung anzuordnen, können Sie auch einfach den folgenden Menübefehl benutzen:

Ausführen > Alles

Wenn Sie anschließend im (designierten) Ausgabefenster keine Spur des Programmlaufs finden, ist alles glatt gegangen. Anderenfalls erscheinen dort Fehlermeldungen und/oder Warnungen in einer mit **Log** betitelten Teilausgabe. Da alle Kommandos Ihres Programms von SPSS erstellt wurden, sollte dies eigentlich nicht passieren.
Ältere Warnungen bzw. Fehlermeldungen sollten vor einem Lauf des Transformationsprogramms aus dem Ausgabefenster gelöscht werden, um Unklarheiten zu vermeiden.

Ein gelungener Lauf des Transformationsprogramms hinterlässt zwar im Ausgabefenster keine Spuren, wirkt sich aber nachhaltig auf das Datenfenster aus. Dort erscheinen z.B. am rechten Rand der Datenmatrix die neuen Variablen.
Sie dürfen aber Ihre Erfolgskontrolle keinesfalls auf das Datenfenster beschränken, sondern müssen unbedingt das Ausgabefenster auf Fehlermeldungen und Warnungen überprüfen. SPSS stoppt nämlich die Programmausführung **nicht** beim Auftreten des ersten fehlerhaften Kommandos, sondern ignoriert das fehlerhafte Kommando und macht unverdrossen mit den nächsten Kommandos weiter. Diese arbeiten aber möglicherweise aufgrund des vorangegangenen Fehlers mit falschen Zwischenergebnissen und produzieren Unsinn. Es kann also leicht passieren, dass nach einem fehlerbehafteten Lauf des Transformationsprogramms z.B. alle erwarteten neuen Variablen vorhanden sind, jedoch unsinnige Werte enthalten.

7.7.3 Übung

Soeben ist mir aufgefallen, dass die Formulierung der Übung 2 in Abschnitt 7.2.3 fehlerhaft ist: Anstatt der LOT-Frage 8 ist die LOT-Frage **5** zu rekodieren. Korrigieren Sie bitte in Ihrem Transformationsprogramm das falsche Kommando, sichern Sie das Programm, und lassen Sie es erneut ausführen.

8 Prüfung der zentralen Projekt-Hypothesen

8.1 Entscheidungsregeln beim Hypothesentesten

In diesem Abschnitt werden einige Grundprinzipien der Inferenzstatistik am Beispiel unserer allgemeinpsychologischen Hypothese demonstriert. Dabei handelt es sich nicht um eine didaktisch sinnvoll ausgearbeitete Einführung, die erheblich mehr Platz beanspruchen würde. Im wesentlichen sollen die statistischen Entscheidungsregeln so präsentiert werden, dass sie mit Hilfe der SPSS-Ausgaben unmittelbar umgesetzt werden können. Zumindest in älteren Statistikbüchern findet man nämlich Formulierungen mit wenig Bezug zu den heute üblichen Ausgaben von Statistikprogrammen.

Wenn mit μ_O der Erwartungswert (Populationsmittelwert) des Merkmals AERGO und mit μ_M der Erwartungswert des Merkmals AERGM bezeichnet wird, dann lautet unser zentrales, allgemeinpsychologisches KFA-Testproblem:

$$\mathrm{H}_0\colon \mu_M \le \mu_O \quad \text{vs.} \quad \mathrm{H}_1\colon \mu_M > \mu_O$$

Mit Hilfe der Differenzvariablen AERGZ := AERGM – AERGO, deren Erwartungswert mit μ_Z bezeichnet werden soll, lässt sich das Testproblem äquivalent noch kompakter formulieren:

$$\mathrm{H}_0\colon \mu_Z \le 0 \quad \text{vs.} \quad \mathrm{H}_1\colon \mu_Z > 0$$

Bei der Reformulierung wird die folgende, generell gültige, Identität ausgenutzt:

$$\mu_Z = \mu_M - \mu_O$$

Wir wollen noch voraussetzen, dass die Differenzvariable AERGZ normalverteilt sei mit dem Erwartungswert μ_Z und der Varianz σ_Z^2:

$$\text{AERGZ} \sim \mathrm{N}(\mu_Z, \sigma_Z^2)$$

Für die *n* AERGZ-Beobachtungen in der Stichprobe nehmen wir an, dass sie durch **unabhängiges** „Ziehen" aus der eben beschriebenen Population entstanden sind. Das schon in Abschnitt 1 betonte Unabhängigkeitsprinzip ist die zentrale Forderung in unserem **Stichprobenmodell** über die Gewinnung der empirischen Daten.

Bei der klassischen inferenzstatistischen Lösung des beschriebenen Testproblems benötigen wir eine sogenannte **Prüfstatistik *T*** mit folgenden Eigenschaften:

1. Die Prüfstatistik *T* kann aus den Stichprobendaten berechnet werden. Sie ist also ein Stichprobenkennwert.
2. Die Prüfstatistik *T* ist indikativ für Abweichungen von der Nullhypothese. Sie wird also tendenziell um so größer, je stärker der Verteilungsparameter μ_Z über den Wert Null hinauswächst.
3. Es ist bekannt, welcher Verteilung der Stichprobenkennwert *T* unter der Nullhypothese folgt, also bei $\mu_Z \le 0$.

In der oben beschriebenen Situation hat sich die folgende Prüfstatistik T_Z bewährt (mit Z als Abkürzung für AERGZ):

$$T_Z := \frac{\bar{Z}}{S_Z}\sqrt{n} \quad \text{mit} \quad \bar{Z} := \frac{1}{n}\sum_{i=1}^{n} Z_i \quad \text{und} \quad S_Z := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2}$$

Diese Prüfstatistik erfüllt obige Anforderungen:

- Sie kann aus den Stichprobendaten berechnet werden (Eigenschaft 1).
- Wir können annehmen, dass mit μ_Z tendenziell auch der Stichprobenmittelwert $\bar{Z}$ ansteigt, der ein erwartungstreuer Schätzer von μ_Z ist. Die geschätzte Standardabweichung S_Z der normalverteilten Variablen Z ist vom Mittelwert μ_Z unabhängig, wächst also bei steigendem μ_Z *nicht* mit. Folglich steigt T_Z mit zunehmender Distanz von der Nullhypothese und erfüllt damit die obige Eigenschaft 2.
- Die restlichen Bestandteile in der Definitionsgleichung zu T_Z sorgen dafür, dass die Prüfstatistik für $\mu_Z = 0$ einer bekannten Verteilung folgt, nämlich der t-Verteilung mit n - 1 Freiheitsgraden. Damit kennen wir das Verhalten der Prüfstatistik für $\mu_Z = 0$, also am „Rand" der Nullhypothese (Eigenschaft 3). Dieses Wissen genügt, weil die bei der Testentscheidung relevante Überschreitungswahrscheinlichkeit unter der H_0 (siehe unten) am Rand der Nullhypothese (also bei $\mu_Z = 0$) maximal wird.

Aufgrund dieser Voraussetzungen kann man den Wert T_{emp} der Prüfstatistik für eine konkrete Stichprobe berechnen und dazu die folgende **Überschreitungswahrscheinlichkeit** bestimmen:

> Mit welcher Wahrscheinlichkeit nimmt die Prüfstatistik T_Z **bei Gültigkeit der Nullhypothese**[1] einen Wert größer oder gleich T_{emp} an?

Diese Wahrscheinlichkeit wollen wir mit $\mathbf{P_{H_0}(T_Z \geq T_{emp})}$ bezeichnen. Sie wird von SPSS berechnet und in der Ausgabe zum t-Test für gepaarte Stichproben mit **Sig.** überschrieben[2].

Bei einem akzeptierten **Fehlerrisiko erster Art** von $\alpha = 5\%$ verwendet man nun die folgende **Entscheidungsregel**:

$$\mathrm{P}_{\mathrm{H}_0}(T_Z \geq T_{\mathrm{emp}}) \begin{cases} \geq 0{,}05 & \Rightarrow \quad \mathrm{H}_0 \text{ beibehalten} \\ < 0{,}05 & \Rightarrow \quad \mathrm{H}_0 \text{ verwerfen} \end{cases} \tag{8-1}$$

Die Nullhypothese wird also abgelehnt, wenn die Prüfstatistik einen Wert annimmt, der bei Gültigkeit der H_0 nur relativ selten (mit einer Wahrscheinlichkeit von < 5%) erreicht oder gar übertroffen wird.

In Statistiklehrbüchern wird oft ein **kritischer Wert** T_{krit} aufgrund der Kenntnis über die Verteilung von T_Z unter der H_0 (genauer: am Rand der H_0) so bestimmt, dass gilt:

$$\mathrm{P}_{\mathrm{H}_0}(T_Z \geq T_{\mathrm{krit}}) = 0{,}05$$

T_{krit} ist gerade das 95%-Quantil der t-Verteilung mit n - 1 Freiheitsgraden.

---

[1] „Bei Gültigkeit der H_0" ist streng genommen zu ersetzen durch „bei $\mu_Z = 0$". Die Überschreitungswahrscheinlichkeit hängt vom Lageparameter μ_Z ab und wird am Rand der Nullhypothese maximal (also bei $\mu_Z = 0$).

2 Leider gibt SPSS beim t-Test für gepaarte Stichproben ausschließlich die zweiseitige Überschreitungswahrscheinlichkeit aus (siehe unten), während wir unsere allgemeinpsychologische KFA-Hypothese mit gutem Grund einseitig formuliert haben und daher auch die einseitige Überschreitungswahrscheinlichkeit $P_{H0}(T_Z \geq T_{emp})$ betrachten.

Bei unserer Stichprobengröße $n = 31$ erhalten wir z.B. $T_{krit} = 1,70$.
Damit kann obige Entscheidungsregel äquivalent folgendermaßen formuliert werden:

$$T_{emp} \begin{cases} \leq T_{krit} \Rightarrow & H_0 \text{ beibehalten} \\ > T_{krit} \Rightarrow & H_0 \text{ verwerfen} \end{cases} \tag{8-2}$$

Wir haben übrigens bei den „quick-and-dirty"-Tests in Abschnitt 5.5 die Testentscheidung anhand von kritischen Werten kennen gelernt. Dort waren wir ausnahmsweise in der Lage, keine Überschreitungswahrscheinlichkeiten zu kennen, aber die kritischen Werte (als Quantile der Standardnormalverteilung) besonders leicht ermitteln zu können.

Wie die Formulierung (8-1) unserer Entscheidungsregel deutlich zeigt, nehmen wir bewusst das Risiko in Kauf, bei Gültigkeit der Nullhypothese mit einer Wahrscheinlichkeit von 5% falsch zu entscheiden (Fehler erster Art, α-Fehler). Der α-Wert sollte um so niedriger angesetzt werden, je gravierender (schädlicher, teurer) das irrtümliche Ablehnen einer gültigen Nullhypothese ist.

Das Risiko, bei Gültigkeit der *Alternativ*hypothese falsch zu entscheiden (**Fehler zweiter Art**, β-Fehler), ist um so kleiner,

- je stärker der wahre Lageparameter μ_Z von der Nullhypothese $\{\mu_Z \leq 0\}$ entfernt ist,
- je größer die Power (Teststärke) unseres Verfahrens ist, d.h. je wahrscheinlicher unter der Alternativhypothese ein „signifikantes" Ergebnis erzielt wird (siehe auch die post hoc Power-Analyse in Abschnitt 8.3.2.2).

Während wir das Risiko erster Art exakt einstellen können, hängt das Risiko zweiter Art u.a. vom unbekannten Verteilungsparameter μ_Z ab und ist daher ebenfalls unbekannt. Für jeden hypothetisch angenommenen Wert von μ_Z kann der β-Fehler aber ausgerechnet werden.

Passend zu unserer allgemeinpsychologischen KFA-Hypothese haben wir bislang das einseitige Testproblem behandelt. Wir wollen noch das folgende **zweiseitige Testproblem** betrachten:

$$H_0\colon \mu_M = \mu_O \quad \text{vs.} \quad H_1\colon \mu_M \neq \mu_O$$

bzw.

$$H_0\colon \mu_Z = 0 \quad \text{vs.} \quad H_1\colon \mu_Z \neq 0$$

Die H_0 des zweiseitigen Tests ist übrigens gerade identisch mit dem *Rand* der Nullhypothese zum einseitigen Test. Die beim einseitigen Testentscheid relevante Überschreitungswahrscheinlichkeit $P_{H_0}(T_Z \geq T_{emp})$ muss zudem nur für diese Randlage ($\mu_Z = 0$) berechnet werden. Wenngleich sich die beiden Nullhypothesen der in diesem Abschnitt betrachteten Testprobleme unterscheiden, ist also bei der Testentscheidung immer nur $P_{\{\mu_Z=0\}}$ relevant.

Wir verwenden beim zweiseitigen Test dieselbe Prüfstatistik T_Z wie beim einseitigen Test. Nun sind aber nicht nur sehr große, sondern auch sehr kleine T_{emp}-Werte indikativ für eine Abweichung von der Nullhypothese. Wie oben schon mitgeteilt wurde, besitzt T_Z bei $\mu_Z = 0$ eine t-Verteilung mit n-1 Freiheitsgraden. Wegen der Symmetrie der t-Verteilung wird man dann sinnvollerweise die folgende Entscheidungsregel wählen:

$$P_{H_0}(|T_Z| \geq |T_{emp}|) \begin{cases} \geq 0{,}05 & \Rightarrow & H_0 \text{ beibehalten} \\ < 0{,}05 & \Rightarrow & H_0 \text{ verwerfen} \end{cases} \tag{8-3}$$

Der kritische Werte $T_{krit,2}$ zum zweiseitigen Test ist so zu bestimmen, dass gilt:

$$P_{H_0}(|T_Z| \geq |T_{krit,2}|) = 0{,}05$$

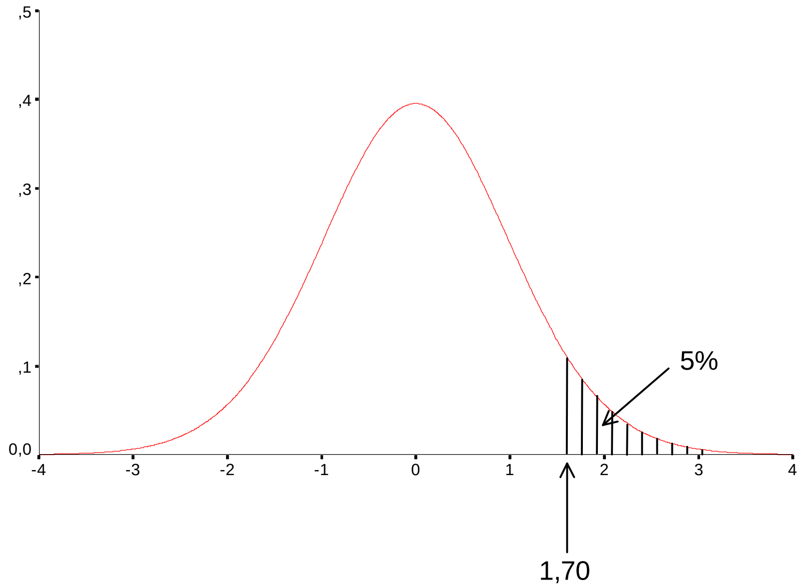
Bei unserer Stichprobengröße $n = 31$ erhalten wir z.B. $T_{krit,2} = \pm 2{,}04$.

Aufgrund der Symmetrie der Prüfverteilung gilt für $T_{emp} \geq 0$:

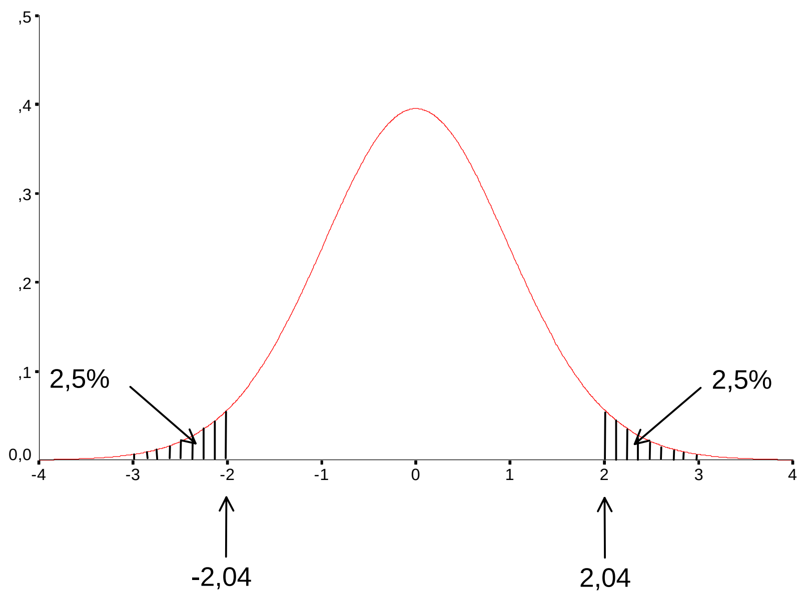
$$P_{H_0}(T_Z \geq T_{emp}) = \frac{1}{2} \cdot P_{H_0}(|T_Z| \geq |T_{emp}|) \tag{8-4}$$

Die Überschreitungswahrscheinlichkeit des einseitigen t-Tests ergibt sich also durch Halbieren aus der Überschreitungswahrscheinlichkeit des zweiseitigen t-Tests. Dieser Zusammenhang ist wichtig in der statistischen Praxis mit SPSS, weil dieses Programm bei t-Tests häufig nur die zweiseitige Überschreitungswahrscheinlichkeit ausgibt. Sie dürfen aber den Zusammenhang in Gleichung (8-4) keinesfalls auf beliebige Tests generalisieren. Wir werden z.B. im Zusammenhang mit der Kreuztabellenanalyse den exakten Test von Fisher kennen lernen, bei dem eine analoge Gleichung **nicht** gilt.

Die folgende Abbildung zeigt unsere Prüfverteilung (t_{30}) sowie den **Ablehnungsbereich** bei einseitiger Fragestellung (|||):



Bei zweiseitiger Fragestellung haben wir zwei symmetrisch angeordnete Ablehnungsbereiche:



8.2 *Verteilungsanalyse zu AERGAM, AERGZ und LOT*

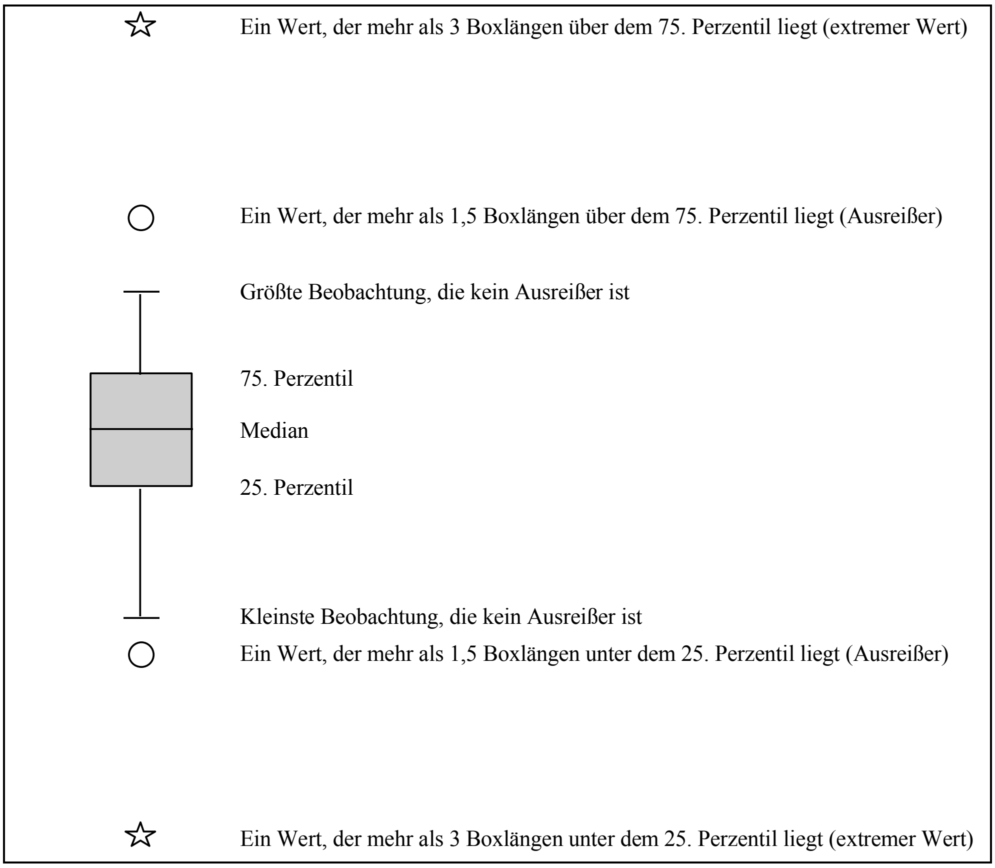
Für die folgenden Schritte wird eine aktive SPSS-Sitzung mit geöffneter Projekt-Fertigdatendatei **kfa.sav** vorausgesetzt. Ob Sie die SPSS-Kommandos zu den durchzuführenden Analysen für spätere Wiederverwendung sichern wollen, bleibt Ihnen überlassen.

Wir wollen zunächst die univariaten Verteilungen der berechneten Variablen AERGAM, AERGZ und LOT untersuchen. Analog zu den Verteilungsanalysen in Abschnitt 5, die auch zur Datenprüfung dienten, wollen wir bei den Verteilungen der abgeleiteten Variablen auch auf Anomalien infolge fehlerhafter oder schlecht durchdachter Berechnungsvorschriften achten. Außerdem wollen wir noch eine weitere Gefahrenquelle für unser Forschungsprojekt ins Visier nehmen:

8.2.1 Diagnose von Ausreißern

Als **Ausreißer** bezeichnet man extreme Werte, die zwar innerhalb des logisch möglichen Wertebereichs liegen, aber doch mit großer Wahrscheinlichkeit nicht aus der interessierenden Verteilung bzw. Population stammen. Diese Werte haben auf viele Auswertungsverfahren einen starken, verzerrenden Einfluss. Daher wollen wir ab jetzt auch auf Ausreißer achten.

Dazu lassen wir uns für jede Variable einen **Boxplot** erstellen. Dieses beliebte Instrument der explorativen Datenanalyse zeigt auf prägnante Weise wesentliche Verteilungs-Informationen, und ist zur Identifikation von Ausreißern sehr gut geeignet. Die Bestandteile eines Boxplots haben folgende Bedeutung:



Als Ursachen für Ausreißer kommen in Frage:

- Erfassungsfehler bzw. Defekte im Messverfahren
  Bei unseren Ärgermessungen in Teil 2 des Fragebogens kann es passieren, dass eine Person die Instruktion nicht richtig liest und die beiden fiktiven Personen verwechselt.
- Besondere Umstände beim Merkmalsträger
  Bei einer Agrarstudie kann z.B. der Boden in einer Versuchsparzelle durch einen Ölunfall verseucht worden sein.

Eindeutig irreguläre Daten müssen natürlich entfernt werden. Sie können z.B. mit dem Dateneditor in der Rohdatendatei:

- einen Wert durch System-Missing ersetzen.
- einen Wert als MD-Indikator deklarieren.
- einen kompletten Fall löschen.

Natürlich dürfen Sie keine Daten eliminieren, weil Sie Ihren Hypothesen widersprechen.

8.2.2 Zu den Normalverteilungsvoraussetzungen unserer Hypothesentests

Im Rahmen der nun anstehenden Verteilungsanalysen sollen auch die Normalverteilungsvoraussetzungen der geplanten Hypothesentests geprüft werden:
Der t-Test für gepaarte Stichproben, mit dem wir unsere allgemeinpsychologische Hypothese prüfen wollen, setzt voraus, dass die Differenzvariable AERGZ normalverteilt ist (vgl. Abschnitt 8.1).
Unsere differentialpsychologische Hypothese soll über die lineare Korrelation $\rho_{A,LOT}$ zwischen AERGAM und LOT geprüft werden:

$$H_0 : \rho_{A,LOT} \geq 0 \quad \text{vs.} \quad H_1 : \rho_{A,LOT} < 0$$

Es kommt die folgende Prüfstatistik zum Einsatz, die auf der Stichprobenkorrelation $r_{A,\,LOT}$ und der Stichprobengröße *n* basiert (siehe z.B. Bortz (1977, Seite 258) oder Hartung (1989, Seite 546)):

$$T_r := \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Sie ist t-verteilt mit *n* - 2 Freiheitsgraden, sofern die beiden Variablen AERGAM und LOT gemeinsam (bivariat) normalverteilt sind. Die Prüfung dieser Voraussetzung macht einige Schwierigkeiten, so dass man sich in der Praxis meist darauf beschränkt, die Normalität der beiden Randverteilungen zu prüfen. Wir werden uns dieser Praxis anschließen.

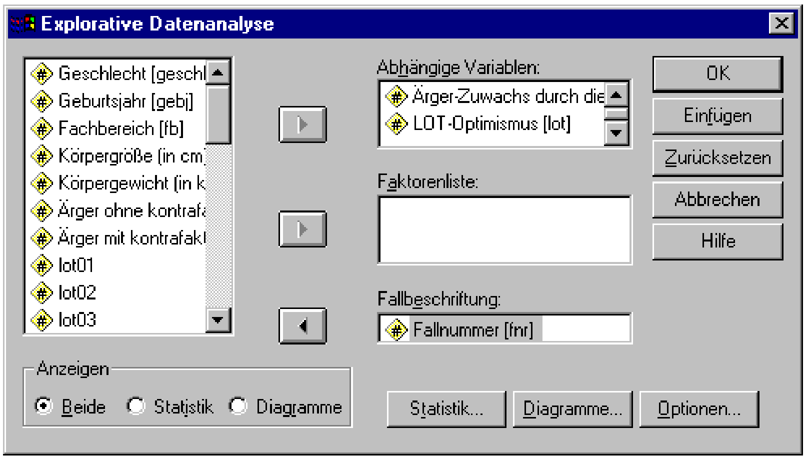
8.2.3 Die SPSS-Prozedur zur explorativen Datenanalyse

Für die eben geplanten Aufgaben (Ausreißerdiagnose und Verteilungsprüfung) eignet sich die SPSS-Prozedur zur explorativen Datenanalyse besser als die in Abschnitt 5 der Einfachheit halber bevorzugte Häufigkeitsanalyse. Natürlich können Sie in Zukunft auch die Verteilungen von Rohvariablen mit der leistungsfähigeren explorativen Datenanalyse untersuchen.

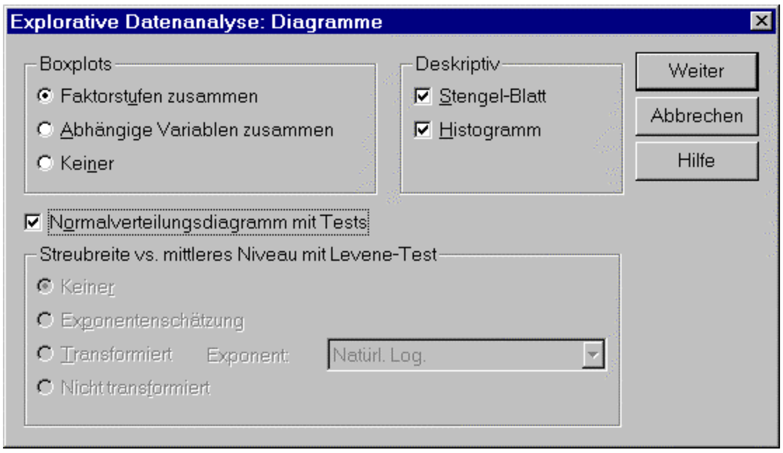
Starten Sie bitte deren Dialogbox mit:

Analysieren > Deskriptive Statistiken > Explorative Datenanalyse...

Transportieren Sie die Namen der drei zu untersuchenden Variablen in die Liste der **abhängigen Variablen** und wählen Sie die Variable FNR zur Fallbeschriftung aus, damit mögliche Ausreißer durch ihre Fallnummer identifiziert werden können:



Fordern Sie in der **Diagramme**-Subdialogbox zusätzlich **Histogramme** sowie **Normalverteilungdiagramme mit Tests** an:

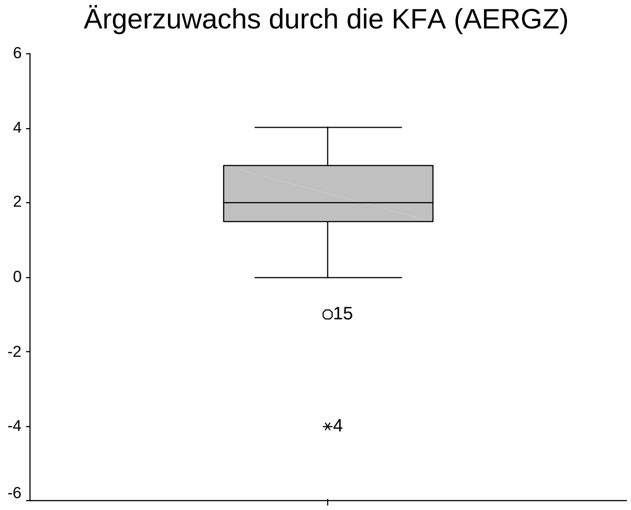


Das Kontrollkästchen zum Anfordern von Normalverteilungs-Anpassungstests (Kolmogorov-Smirnov und Shapiro-Wilk) hat SPSS wirklich sehr gut in der **Diagramme**-Subdialogbox der explorativen Datenanalyse versteckt.

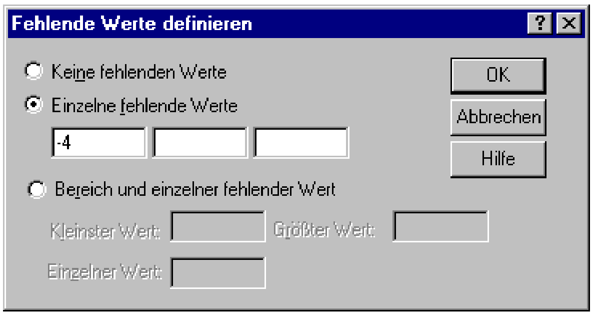
Wir erhalten im Viewer-Fenster u.a. für jede abhängige Variable einen **Boxplot**.

8.2.4 Ergebnisse für AERGZ

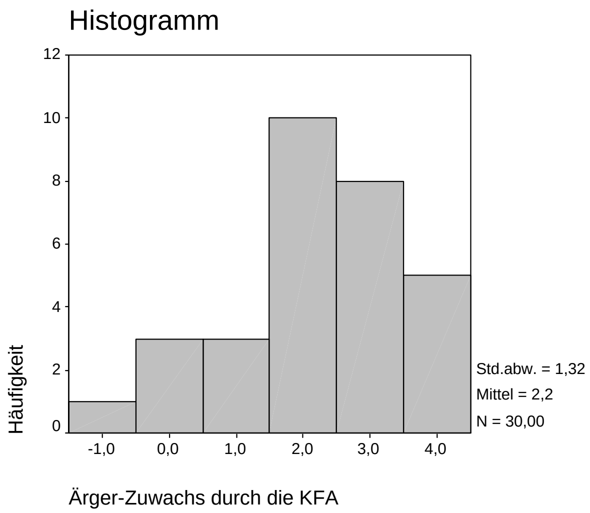
Bei der Ausreißer-Analyse gibt es nur einen Problemfall und zwar ausgerechnet bei der Variablen AERGZ, über die unsere zentrale KFA-Hypothese geprüft werden soll. Hier tanzt Fall Nr. 4 aus der Reihe:



Diese Person hatte *ohne* KFA eine Ärgertemperatur von 60° gemeldet, die sich dann durch die KFA-Komponente angeblich auf 20° abkühlte. Das Muster erscheint sehr unplausibel, und der Boxplot gibt eine klare Empfehlung, den Fall auszuschließen. Allerdings scheut sich ein redlicher Forscher, Daten zu neutralisieren, die der eigenen Hypothese widersprechen.
Vor einer endgültigen Entscheidung wollen wir die Verteilung von AERGZ noch weiter analysieren, da beim geplanten t-Test zur allgemeinpsychologischen KFA-Hypothese vorausgesetzt werden muss, dass AERGZ (in der Population) normalverteilt ist.
Damit der extreme AERGZ-Wert von Fall Nr. 4 die weitere Verteilungsanalyse nicht beeinflusst, soll er vorübergehend neutralisiert werden. Weil wir noch keine Methode kennen, komplette Fälle von einer Analyse fern zu halten (siehe Abschnitt 10), deklarieren wir den betroffenen Wert (= -4) als MD-Indikator. Auf diese Weise finden sich doch noch eine Gelegenheit, die Deklaration von benutzerdefinierten MD-Indikatoren zu üben. Markieren Sie in der Variablenansicht des Datenfensters die Zelle mit den **Fehlenden Werten** der Variablen AERGZ, klicken Sie auf den Erweiterungsschalter ···, und tragen Sie den Wert –4 als einzelnen MD-Indikator ein:



Das folgende Histogramm zeigt, dass die AERGZ-Verteilung *auch nach* Elimination von Fall Nr. 4 noch relativ deutlich von der Normalität abweicht:



Tatsächlich lehnen auch *nach* der Elimination des Ausreißers die beiden von SPSS angebotenen Normalverteilungstests (Kolmogorov-Smirnov und Shapiro-Wilk) die im t-Test benötigte Normalverteilungsannahme ab:

Tests auf Normalverteilung

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Ärger-Zuwachs durch die KFA	,207	30	,002	,911	30	,019

a. Signifikanzkorrektur nach Lilliefors

Auch diese Testentscheidung folgt der in Abschnitt 8.1 beschriebenen Logik, wobei folgende Hypothesen zur Konkurrenz stehen:

H_0: AERGZ ist normalverteilt versus H_1: AERGZ ist nicht normalverteilt

Die von SPSS berechnete Überschreitungswahrscheinlichkeit (**Signifikanz**) ist bei beiden Prüfstatistiken kleiner als 5%, so dass beide Tests übereinstimmend die Nullhypothese verwerfen. Dies ist vor allem deshalb ein ernst zu nehmender Befund, weil unsere Stichprobe relativ klein und damit die Power der Tests eher gering ist.
Bei einer *großen* Stichprobe besitzen die Normalitätstests eine hohe Power und decken auch kleinste (für den t-Test irrelevante) Abweichungen von der Nullhypothese auf. Folglich ist dann ein signifikantes Testergebnis „nicht tragisch“. Wenn bei einer *kleinen* Stichprobe ein Normalitätstest „anschlägt“, muss jedoch die Verteilungs-Pathologie schon ernsthaft sein.

Aufgrund der problematischen Verteilungsverhältnisse entscheiden wir uns, statt des geplanten (parametrischen) t-Tests für gepaarte Stichproben, einen verteilungsfreien Lokationsvergleich mit dem **Vorzeichentest** durchzuführen (siehe z.B. Hartung 1989, S. 242f).
Dieser Test entscheidet sich zwischen folgenden Hypothesen:

H_0: Der Median der Differenzvariablen AERGZ ist kleiner oder gleich Null.

versus

H_1: Die Differenzvariable AERGZ hat einen positiven Median.

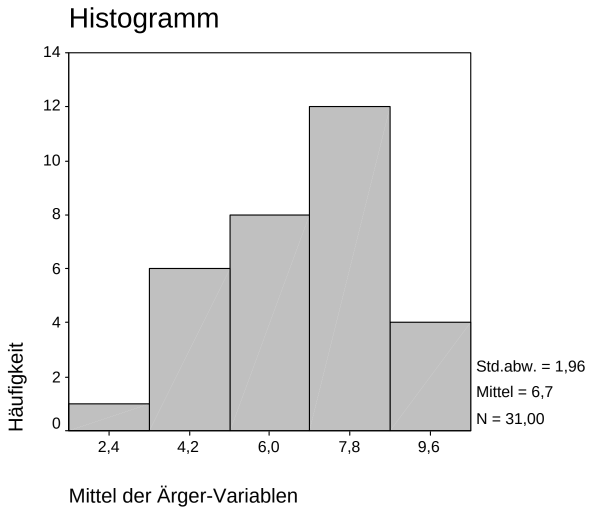
Statt der in Abschnitt 8.1 ausführlich vorgestellten Prüfstatistik T_Z verwendet der Vorzeichentest eine Prüfgröße, die im wesentlichen auf der Anzahl der positiven AERGZ-Ausprägungen in der Stichprobe basiert. Sie wird üblicherweise mit *Z* bezeichnet, weil sie unter der H_0 (genauer: bei einem Median von Null) approximativ z- verteilt (d.h. standardnormalverteilt) ist. Die Übereinstimmung der Bezeichnung mit der oben eingeführten Abkürzung für unsere Ärgerzuwachs-Variable ist also rein zufällig.
Man geht davon aus, dass die Verteilungs-Approximation ab $n \geq 20$ hinreichend genau ist, so dass wir den Test bei unserer Stichprobe ($n = 31$) in der üblichen approximativen Form anwenden dürfen. Bei kleineren Stichproben muss eine exakte Variante des Tests eingesetzt werden, die von SPSS ebenfalls unterstützt wird (siehe unten).

Weil der Vorzeichentest weit weniger empfindlich auf Ausreißer reagiert als der parametrische t-Test, können wir den kritischen Fall Nr. 4 in der Auswertung belassen. Damit vermeiden wir den Verdacht, die Daten zu unseren Gunsten bereinigt zu haben.
Heben Sie also bitte die MD-Deklaration für den Wert –4 bei der Variablen AERGZ wieder auf.

Die bisherige Diskussion der AERGZ-Verteilung hat sich auf Gefahrenquellen für die Interpretierbarkeit des geplanten zentralen Hypothesentests konzentriert. Es ist jedoch keinesfalls verboten, sondern sogar dringend empfohlen, sich anhand obiger Verteilungsdiagramme und sonstiger deskriptiver Informationen einen Eindruck von der empirischen Bewährung der Hypothese zu verschaffen. Das Histogramm spricht für einen starken KFA-Effekt in der erwarteten Richtung. Eine genaue Kenntnis des deskriptiven Ergebnisbildes kann verhindern, dass wir uns von einem durch technische Defekte verfälschten Testergebnis in die Irre führen lassen.

8.2.5 Ergebnisse für AERGAM

Bei der Summenvariablen AERGAM wirkt sich der, wie wir inzwischen wissen, deutlich schief verteilte „Summand" AERGM nicht sehr dramatisch aus:



Die von SPSS berechneten Normalitätstests (Shapiro-Wilk und K-S) akzeptieren beide die Nullhypothese der Normalverteilung (Überschreitungswahrscheinlichkeit jeweils größer als 5%):

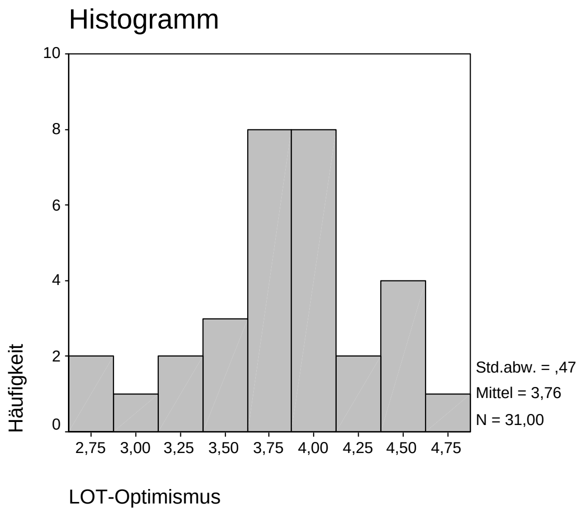
Tests auf Normalverteilung

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Mittel der Ärger-Variablen	,141	31	,117	,947	31	,199

a. Signifikanzkorrektur nach Lilliefors

8.2.6 Ergebnisse für den LOT

Die LOT-Verteilung macht einen soliden Eindruck:



Die Normalitäts-Nullhypothese wird von beiden Tests beibehalten:

Tests auf Normalverteilung

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
LOT-Optimismus	,127	31	,200*	,961	31	,413

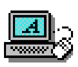
*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

Da sich AERGAM und LOT im univariaten Normalverteilungstest bewährt haben, bestehen von daher keine Einwände gegen die Prüfung der differentialpsychologischen Hypothese über eine lineare Korrelationsanalyse.

8.3 Prüfung der differentialpsychologischen Hypothese

8.3.1 Korrelation zwischen LOT und AERGAM



Nun wollen wir die lineare Beziehung zwischen der Optimismus-Variablen LOT und der Ärgervariablen AERGAM überprüfen, indem wir ihre Produkt-Moment- (alias Pearson-) Korrelation berechnen lassen. Die zuständige Dialogbox öffnet sich nach

Analysieren > Korrelation > Bivariat...

Da wir eine gerichtete Hypothese prüfen, müssen wir im Rahmen **Test auf Signifikanz** statt des voreingestellten zweiseitigen Tests den einseitigen wählen:



Wir erhalten zwar, wie erwartet, eine negative Korrelation, doch ist diese sehr schwach ausgeprägt und bei weitem nicht signifikant. Die Überschreitungswahrscheinlichkeit (Bez.: **Signifikanz (1-seitig)**) ist erheblich größer als 0,05:

Korrelationen

		Mittel der Ärger-Variablen	LOT-Optimismus
Mittel der Ärger-Variablen	Korrelation nach Pearson	1,000	-,063
	Signifikanz (1-seitig)	,	,369
	N	31	31
LOT-Optimismus	Korrelation nach Pearson	-,063	1,000
	Signifikanz (1-seitig)	,369	,
	N	31	31

Der Signifikanztest entscheidet sich sehr deutlich für die Nullhypothese, also gegen unsere differentialpsychologische Alternativhypothese. Der LOT-Optimismus zeigt entgegen unserer Annahme fast keinen *linearen* Zusammenhang mit dem summativen Ärger in unserer fiktiven Situation.

8.3.2 Methodologische Anmerkungen

8.3.2.1 Explorative Analysen im Anschluss an einen „gescheiterten" Hypothesentest

Nach der Pleite einer konfirmatorischen Forschungsbemühung wird sich in der Regel eine exploratorische Phase anschließen. Im Fall unserer differentialpsychologischen Hypothese sollten wir uns spätestens jetzt mit Hilfe eines Streuungsdiagramms (siehe unten) einen Eindruck von der bivariaten Verteilung der beiden Variablen verschaffen. Oben wurde schon zu Recht festgestellt, dass man (wegen potentieller technischer Defekte) einem statistischen Test nur dann glauben sollte, wenn seine Entscheidung mit den deskriptiven Befunden harmoniert. Wir mussten bislang auf das Streuungsdiagramm verzichten, weil uns die dazu nötigen SPSS-Kenntnisse noch fehlen.

Außer dem Streuungsdiagramm kommen in unserem Beispiel auch noch andere statistische und graphische Analysen in Frage, um neue Information über empirische Gesetzmäßigkeiten zu gewinnen. Bei der explorativen Analyse der Stichprobendaten können Hypothesen generiert oder verbessert werden. Wir werden uns in Abschnitt 9 z.B. dafür interessieren, ob eventuell das Geschlecht den Zusammenhang zwischen Optimismus und Ärger moderiert. Allerdings ist es unmöglich, die revidierten Hypothesen anhand *derselben* Stichprobe zu testen. Also: Sie dürfen

und sollen aus Ihren Daten etwas lernen, aber ein echter Test einer dabei generierten Hypothese ist nur in einer neuen, unabhängigen Stichprobe möglich.

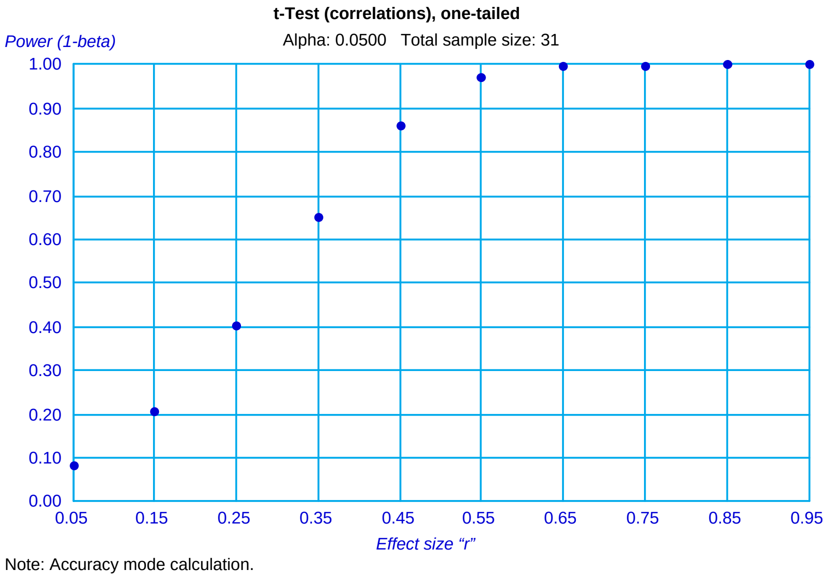
Außerdem sollten Sie es nicht unterlassen, das „Scheitern" einer Hypothese zu veröffentlichen. Ansonsten tragen Sie dazu bei, in der Fachliteratur ein systematisch verzerrtes Bild der Wirklichkeit aufzubauen.

8.3.2.2 Post hoc - Poweranalyse

Bei der Interpretation des obigen Resultates ist außerdem zu beachten, dass die Power des t-Test in unserer relativ kleinen Stichprobe nicht sehr gut ist. Z.B. beträgt die Wahrscheinlichkeit, bei einer wahren Korrelation von $\rho = 0{,}2$ in einer Stichprobe der Größe $n = 31$ mit dem einseitigen t-Test für Pearson-Korrelationen ein signifikantes Ergebnis zu finden nur:

$$\text{Power(t-Test; } \rho = 0{,}2;\ n = 31) = 0{,}2964$$

Die folgende Abbildung zeigt, wie bei gegebener Stichprobengröße ($n = 31$) die Power des einseitigen t-Tests von der Effektstärke, d.h. von der wahren Korrelation ρ abhängt:



Sie wurde mit dem exzellenten Power-Analyse-Programm **Gpower** erzeugt (Erdfelder, Faul & Buchner 1996), das Sie kostenlos über folgende Webseite beziehen können:

http://www.psychologie.uni-trier.de:8000/projects/gpower.html

8.3.2.3 Paarweiser oder fallweiser Ausschluss fehlender Werte

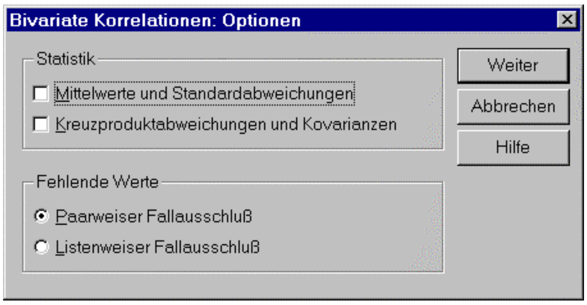
Wir müssen uns leider wieder einmal mit dem Problem fehlender Werte befassen, wenn auch ohne direkten Bezug auf unser Demoprojekt: Wenn Sie die Korrelationsmatrix zu gewissen Variablen A, B, C und D anfordern, dann kann SPSS fehlende Werte auf zweierlei Weise berücksichtigen:

- **Paarweiser Ausschluss fehlender Werte**
  Zur Berechnung der Korrelation zwischen den Variablen A und B werden alle Fälle herangezogen, die *bei diesen beiden* Variablen einen validen Wert haben.
  Vorteil: Alle verfügbaren validen Beobachtungen werden ausgenutzt.
  Nachteil: In der entstehenden Korrelationsmatrix beruhen die einzelnen Koeffizienten im allgemeinen auf unterschiedlichen Teilstichproben. Daher fehlt dieser Matrix eine gewisse mathematische Eigenschaft (die positive Se-

midefinitheit), die bei normalen Korrelationsmatrizen vorhanden ist und die in vielen Statistikprozeduren vorausgesetzt wird. Es kann dadurch (z.B. in einer multiplen Regressionsanalyse) zu artifiziellen Ergebnissen kommen.

- **Fallweiser Ausschluss fehlender Werte**
  Zur Berechnung der Korrelation zwischen den Variablen A und B werden nur Fälle herangezogen, die *bei allen Variablen*, also bei A, B, C und D, einen validen Wert haben.
  Vorteil: Die entstehende Korrelationsmatrix ist intakt (positiv semidefinit).
  Nachteil: Ist die Gesamtmenge der beteiligten Variablen groß, fallen eventuell sehr viele Fälle aus.

Per Voreinstellung benutzt SPSS bei der Korrelationsberechnung die *paarweise* Methode. Mit dem Schalter **Optionen** in der Dialogbox **Bivariate Korrelationen** erhalten Sie folgende Subdialogbox, die ein Umschalten auf die fallweise Methode erlaubt:



Bei anderen Statistik-Prozeduren kann SPSS durchaus die alternative Voreinstellung „fallweise" benutzen.

Auf die diversen Methoden, fehlende Werte durch Schätzungen zu ersetzen, die in SPSS über

Transformieren > Fehlende Werte ersetzen...

erreichbar sind, kann an dieser Stelle nicht eingegangen werden.

8.4 Prüfung der KFA-Hypothese

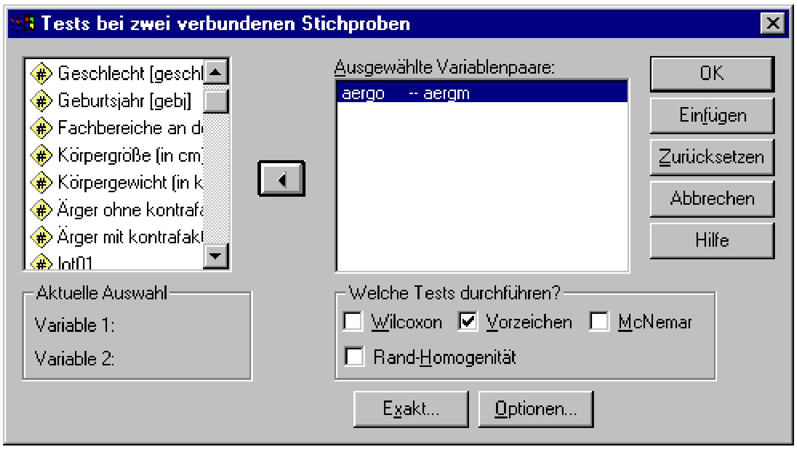
Nun wollen wir die allgemeinpsychologische Ausgangshypothese unserer Studie prüfen, dass die Verfügbarkeit kontrafaktischer (also positiver) Alternativen den Ärger über ein ungünstiges Ereignis steigert. Aufgrund der Ausreißer- und Verteilungsanalyse in Abschnitt 8.2.4 haben wir uns entschieden, statt des ursprünglich geplanten (parametrischen) t-Tests für abhängige Stichproben den verteilungsfreien Vorzeichentest zu verwenden.
Suchen Sie die zuständige Dialogbox zunächst über das **Analysieren**-Menü. Bei Misserfolg können Sie auch den Index des Hilfesystems benutzten. Steigen Sie ein mit:

Hilfe > Themen > Index

und beginnen Sie dann, in das aktive Textfeld „Vorzeichentest" zu schreiben. Schon nach dem vierten Buchstaben erscheint in der Stichwortliste u.a. der etwas verunglückte Eintrag **Vorzeichentext:bei Tests bei zwei verbundenen Stichproben**, den Sie per Doppelklick wählen können. Nach einem weiteren Mausklick auf den Schalter **Verfahren** wird u.a. der Weg zur benötigten Dialogbox erklärt:

Analysieren > Nichtparametrische Tests > Zwei verbundene Stichproben

In der Dialogbox müssen Sie die beiden Variablen angeben und den gewünschten Test markieren:



Wir erhalten folgendes Ergebnis:

Häufigkeiten

		N
Ärger mit kontrafaktischer Alternative - Ärger ohne kontrafaktische Alternative	Negative Differenzen[a]	2
	Positive Differenzen[b]	26
	Bindungen[c]	3
	Gesamt	31

a. Ärger mit kontrafaktischer Alternative < Ärger ohne kontrafaktische Alternative

b. Ärger mit kontrafaktischer Alternative > Ärger ohne kontrafaktische Alternative

c. Ärger ohne kontrafaktische Alternative = Ärger mit kontrafaktischer Alternative

Statistik für Test[a]

	Ärger mit kontrafaktischer Alternative - Ärger ohne kontrafaktische Alternative
Z	-4,347
Asymptotische Signifikanz (2-seitig)	,000

a. Vorzeichentest

Selbst die von SPSS ausgegebene *zweiseitige* Überschreitungswahrscheinlichkeit (Bezeichnung: **Signifikanz (2-seitig)**) ist deutlich kleiner als unser vorgegebenes α-Niveau (0,05). Das unserer einseitigen Fragestellung entsprechende *einseitige* p-level ergibt sich (wegen der Symmetrie der zugrundeliegenden Prüfverteilung) durch Halbierung des zweiseitigen p-levels, ist also erst recht kleiner als die kritische Grenze 0,05.
Damit kann die allgemeinpsychologische KFA-Nullhypothese („Kein Ärgerzuwachs durch eine kontrafaktische Alternative“) deutlich zurückgewiesen werden.

Nach Klärung der zentralen Hypothesen ist unser Projekt nun eigentlich abgeschlossen, aber es gibt noch viele SPSS-Optionen kennen zu lernen, und unsere Daten enthalten sicher auch noch einige interessante Details.

8.5 Übung

Für die Differenzvariable (GEWICHT - IDGEW) akzeptiert der Shapiro-Wilk-Test die Normalverteilungshypothese:

Test auf Normalverteilung

	Shapiro-Wilk		
	Statistik	df	Signifikanz
IDGEW - GEWICHT	,984	31	,917

Führen Sie bitte mit den Variablen GEWICHT und IDGEW einen t-Test für gepaarte Stichproben zu folgendem Testproblem durch:

H_0: Das Realgewicht der Trierer Studierenden liegt im Mittel nicht unter dem Idealgewicht nach der Formel „Größe - 100".

versus

H_1: Die Trierer Studierenden sind in Relation zur Idealgewichtsformel „Größe - 100" im Mittel zu leicht.

Die Ergebnisse werden im nächsten Abschnitt wiedergegeben.

8.6 Arbeiten mit dem Ausgabefenster (Teil III)

8.6.1 Der Pivot-Editor

Oben wurde gelegentlich in didaktischer Nachlässigkeit ohne Erläuterung der Begriff *Pivot-Tabelle* verwendet. Unter dem *Pivotieren* einer Tabelle versteht SPSS u.a. die folgenden Operationen:

- Austauschen ihrer Zeilen-, Spalten- und Schichtendimensionen
- Änderung der Schachtelungsordnung
- Verstecken von Zellen

Nachdem wir den Pivot-Editor im zweiten Teil der Serie *Arbeiten mit dem Ausgabefenster* bereits für „konventionelle" Tabellengestaltungen benutzt haben, beschäftigen wir uns nun mit den Leistungen, die seinen Namen begründen.

8.6.1.1 Pivot-Editor starten

Man startet den Pivot-Editor zum **Bearbeiten** einer Tabelle per Mausdoppelklick oder über das Kontextmenü. Es empfiehlt sich, anschließend nötigenfalls mit dem Menübefehl **Pivot > Pivot-Leisten** das folgende Fenster einzuschalten:



Es enthält je eine Leiste für die Zeilen, Spalten und Schichten der Tabelle und je ein Pivotsymbol  für die dargestellten Tabellendimensionen. Die Zeilenleiste enthält z.B. die Pivotsymbole

zu allen in den Zeilen dargestellten Tabellendimensionen. Welche Dimension ein Symbol repräsentiert, erfährt man per PopUp-Text, wenn man den Mauszeiger einige Zeit darauf ruhen lässt.
Wir wollen als Beispiel die in obiger Übung von Ihnen erstellte Tabelle mit dem t-Test zum Vergleich von Real- und Idealgewicht betrachten:

Test bei gepaarten Stichproben

		Gepaarte Differenzen							
					95% Konfidenzintervall der Differenz				
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	Untere	Obere	T	df	Sig. (2-seitig)
Paaren 1	Körpergewicht (in kg) - Idealgewicht nach der Formel: Größe - 100	-9,3226	6,1881	1,1114	-11,5924	-7,0528	-8,388	30	,000

Diese Tabelle enthält leider nur eine Schicht, so dass wir den Umgang mit Mehrschichttabellen nicht üben können. Die Zeilenleiste unserer Tabelle zeigt **Paare** als einzige Dimension. Da wir nur ein einziges Variablenpaar untersucht haben, hat diese Dimension zudem nur eine einzige Kategorie, deren Beschriftung aus den Labels der beiden Variablen abgeleitet wurde. Als Kategorien der Spaltendimension **Statistik** fungieren die angezeigten Auswertungsfunktionen (z.B. Mittelwert, Standardabweichung).

8.6.1.2 Dimensionen verschieben

Durch das Verschieben ihres Pivotsymbols kann man für eine Dimension neu festlegen, ob ihre Kategorien durch Spalten, Zeilen oder Schichten dargestellt werden sollen. Wenn in unserem Beispiel die beiden Pivotsymbole ihre Plätze tauschen, benötigt die Tabelle in horizontaler Richtung deutlich weniger Platz:

Test bei gepaarten Stichproben

			Paaren 1
			Körpergewicht (in kg) - Idealgewicht nach der Formel: Größe - 100
Gepaarte Differenzen	Mittelwert		-9,3226
	Standardabweichung		6,1881
	Standardfehler des Mittelwertes		1,1114
	95% Konfidenzintervall der Differenz	Untere	-11,5924
		Obere	-7,0528
T			-8,388
df			30
Sig. (2-seitig)			,000

8.6.1.3 Gruppierungen

Man kann mehrere Kategorien einer Dimension zusammenfassen und mit einem Gruppenetikett kennzeichnen. In der aktuellen Version unserer Beispieltabelle sind z.B. die **Untere** und die **Obere** Konfidenzschranke gruppiert mit dem Etikett **95% Konfidenzintervall der Differenz**. Beseitigen Sie bitte diese Gruppierung folgendermaßen:

- Rechtsklick auf das Kategorienetikett
- Aus dem Kontextmenü wählen: **Gruppierung aufheben**

Welche Gruppierungen die SPSS-Programmierer zur Gestaltung einer Tabellen eingebaut haben, erkennt man am besten nach dem Einschalten der Gitterlinien mit

Ansicht > Gitterlinien

In unserem Beispiel zeigt sich bei der Statistikdimension eine weitere Gruppe mit dem Etikett **Gepaarte Differenzen**:[1]

Test bei gepaarten Stichproben

		Paaren 1
		Körpergewicht (in kg) - Idealgewicht nach der Formel: Größe - 100
Gepaarte Differenzen	Mittelwert	-9,3226
	Standardabweichung	6,1881
	Standardfehler des	1,1114
	Untere	-11,5924
	Obere	-7,0528
T		-8,388
df		30
Sig. (2-seitig)		,000

Beseitigen Sie bitte der Übersichtlichkeit halber auch diese Gruppierung.
Wenn Sie schließlich noch bei der **Paare**-Dimension das Gruppenetikett **Paaren 1** entfernen, erhalten Sie folgendes Zwischenergebnis:

Test bei gepaarten Stichproben

	Körpergewicht (in kg) - Idealgewicht nach der Formel: Größe - 100
Mittelwert	-9,3226
Standardabweichung	6,1881
Standardfehler des	1,1114
Untere	-11,5924
Obere	-7,0528
T	-8,388
df	30
Sig. (2-seitig)	,000

Wenn Sie mehrere Kategorien einer Dimension zu einer Gruppe zusammenfassen wollen, können Sie folgendermaßen vorgehen:

- Alle Kategorien markieren
- Kontextmenü zu einer markierten Kategorie öffnen und Option **Gruppieren** wählen
- Gruppenbeschriftung anpassen

In der folgenden Version unserer Tabelle wurde eine Gruppe mit den 3 Kategorien zum t-Test gebildet:

[1] Eingeblendete Gitterlinien sind nur bei aktivem Pivot-Editor sichtbar. Um diese Hilfslinien im Manuskript darzustellen, wurden über **Format > Tabelleneigenschaften > Rahmen** zusätzliche Trennlinien aktiviert (und später wieder abgeschaltet).

Test bei gepaarten Stichproben

		Körpergewicht (in kg) - Idealgewicht nach der Formel: Größe - 100
Mittelwert		-9,32
Standardabweichung		6,19
Standardfehler des Mittelwertes		1,11
Untere		-11,59
Obere		-7,05
Signifikanztest	T	-8,39
	df	30
	Sig. (2-seitig)	,00

Außerdem wurde bei einigen Zellen die Anzahl der Dezimalstellen reduziert (über **Format > Zelleneigenschaften**).

8.6.1.4 Kategorien aus- und einblenden

Wenn eine SPSS-Tabelle zu ausführlich erscheint, können Kategorien einer Dimension ausgeblendet werden. In unserem Beispiel wollen wir bei der Statistikdimension auf den Standardfehler des Mittelwertes und die Konfidenzintervalle verzichten:

Test bei gepaarten Stichproben

		Körpergewicht (in kg) - Idealgewicht nach der Formel: Größe - 100
Mittelwert		-9,32
Standardabweichung		6,19
Signifikanztest	T	-8,39
	df	30
	Sig. (2-seitig)	,000

Gehen Sie beim Ausblenden einer Kategorie folgendermaßen vor:

- Setzen Sie bei gedrückter Tastenkombination **<Strg>+<Alt>** einen (linken) Mausklick auf das Kategorienetikett.
- Rechtsklick auf das Kategorienetikett
- Aus dem Kontextmenü wählen: **Kategorie ausblenden**

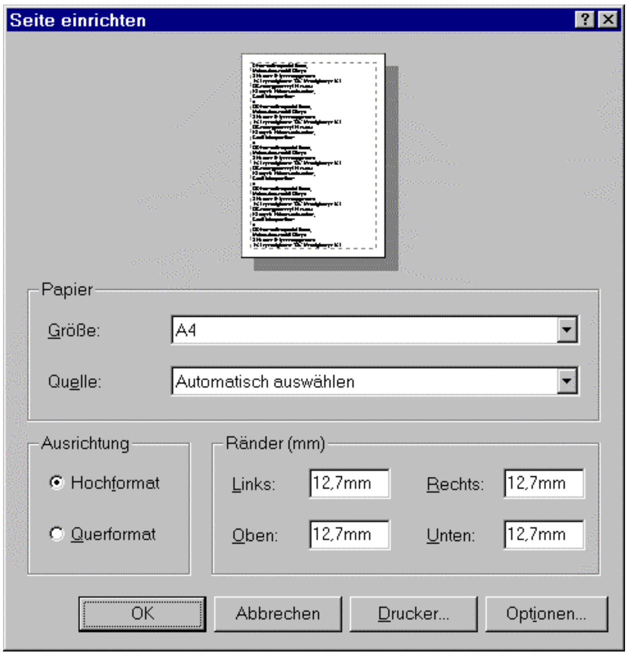
Zum *Einblenden* von vorher abgeschalteten Kategorien kenne ich nur die global wirksame Methode:

Ansicht > Alles einblenden

Nach diesem Befehl können Tabellenbestandteile auftauchen (z.B. Dimensionsbeschriftungen), die (je nach verwendeter Vorlage) bei neuen Tabellen nicht eingeschaltet sind.

8.6.2 Druckausgabe des Viewer-Fensters optimieren

Beim Ausdrucken des Viewer-Fensters stehen über **Datei > Seite einrichten** einige Gestaltungsmöglichkeiten zur Verfügung:



In der **Optionen**-Subdialogbox können u.a. Kopf- und Fußzeilen festgelegt werden, z.B. unter Verwendung von Standardelementen wie **Überschrift erster Stufe** , **Datum** und **Uhrzeit** .
Den Erfolg Ihrer Bemühungen können Sie über **Datei > Seitenansicht** auch schon vor dem Ausdrucken begutachten.

8.6.3 Der Text-Viewer

Wer SPSS-Ergebnisse mit einem Schnelldrucker zu Papier bringen möchte, der auf ASCII-Zeichen beschränkt ist, kann an Stelle der Pivot-Tabellen textorientierte Ausgaben erzeugen lassen.
Zum Umschalten auf textorientierte Ausgaben muss man lediglich mit

Datei > Neu > Textausgabe

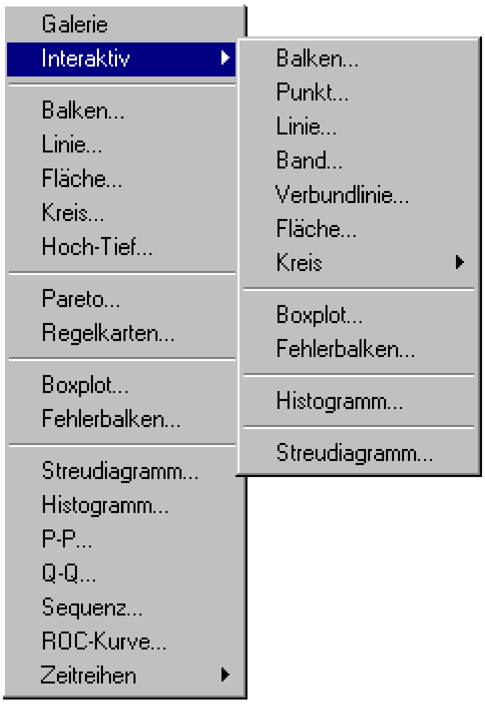
ein passendes Ausgabefenster anfordern. SPSS betrachtet das vom sogenannten Text-Viewer verwaltete Fenster automatisch als designiertes Ausgabeziel und produziert dementsprechend textorientierte Ergebnisse.
Wer permanent umsteigen will, sollte nach dem Menübefehl **Bearbeiten > Optionen** auf dem Registerblatt **Allgemein** der Dialogbox **Optionen** im Bereich **Ausgabetyp beim Starten** die Voreinstellung **Viewer** durch die Alternative **Text-Viewer** zu ersetzen.
Die Tabelle mit der Korrelation von AERGAM und LOT sieht im Text-Viewer folgendermaßen aus:

		Mittel der Ärger-Variablen	LOT-Optimismus
Mittel der Ärger-Variablen	Korrelation nach Pearson	1,000	-,063
	Signifikanz (1-seitig)	,	,369
	N	31	31
LOT-Optimismus	Korrelation nach Pearson	-,063	1,000
	Signifikanz (1-seitig)	,369	,
	N	31	31

Ein Pivotieren ist bei den Tabellen im Text-Viewer nicht möglich.

9 Graphische Datenanalyse

SPSS kann zahlreiche Grafiktypen produzieren und erlaubt darüber hinaus eine individuelle Nachbearbeitung. SPSS-Einsteiger werden vermutlich durch das **Grafiken**-Menü leicht irritiert, weil viele Grafiktypen sowohl auf der Hauptebene als auch im Untermenü **Interaktiv** auftauchen:



Ursache ist die aus Kompatibilitätsgründen erforderliche Koexistenz der klassischen SPSS-Grafik (verknüpft mit dem Kommando GRAPH) mit der moderneren interaktiven Grafik (verknüpft mit dem Kommando IGRAPH).
Die von uns in diversen Statistik-Dialogboxen zahlreich angeforderten Grafiken werden noch über das GRAPH-Kommando erzeugt. In diesem Abschnitt wird die interaktive Grafik vorgestellt, uns Zeitgründen beschränkt auf wenige Beispiele.

Relativ ausführlich wird das Streuungsdiagramm zur Veranschaulichung der gemeinsamen Verteilung zweier Variablen behandelt. Als Anwendungsbeispiel betrachten wir die Regression von Größe auf Gewicht. Bei der Übung in Abschnitt 8.5 zum Vergleich der realen Gewichtsvariablen mit dem Idealgewicht nach der Formel "Größe - 100" sollten Sie folgende Ergebnisse erhalten haben:

Mittelwertsdifferenz:	-9,3226
t-Wert:	-8,39
Überschreitungswahrscheinlichkeit:	0,00006

Die Trierer Studierenden in unserer Stichprobe wiegen also im Schnitt ca. 9,3 kg weniger, als die *normative* Idealgewichtsformel erlaubt. Die Überschreitungswahrscheinlichkeit für den ermittelten t-Wert von -8,39 ist deutlich kleiner als 0,05, so dass die Nullhypothese klar verworfen wird.[1]

[1] Wie Sie bereits wissen, teilt SPSS beim t-Test nur die *zweiseitige* Überschreitungswahrscheinlichkeit mit. Beim Testen einer *einseitigen* Hypothese muss man daher die ausgegebene Wahrscheinlichkeit halbieren.
Per Voreinstellung gibt SPSS die Überschreitungswahrscheinlichkeit mit drei Dezimalstellen aus, wobei man über den Pivot-Editor auch den exakten Wert ermitteln kann (vgl. Abschnitt 5.7).

Da die normative Idealgewichtsformel „Größe - 100“ (als Regressionsgleichung aufgefasst) offenbar die Verhältnisse in der untersuchten Population nicht gut beschreibt, wollen wir die tatsächliche empirische Regression von Gewicht auf Größe analysieren, wobei wir den Regressionstyp (z.B. linear, logarithmisch) in explorativer Manier aus dem bivariaten Streuungsdiagramm ablesen wollen.

9.1 Erstellung eines interaktiven Streudiagramms über das Grafiken-Menü

Nach

Grafiken > Interaktiv > Streudiagramm

erscheint die Dialogbox **Streudiagramm erstellen**, wobei das Registerblatt **Variablen zuweisen** im Vordergrund liegt.

0.1.1 Variablen zuweisen

Bei den Dialogboxen zu den interaktiven Diagrammen hat SPSS einige neue Bedienungsmerkmale realisiert.

9.1.1.1 Manipulation der Anwärterliste

Die Änwärter-Variablenliste weist gegenüber den SPSS-Dialogboxen zur Spezifikation von statistischen Analysen bzw. von Standarddiagrammen einige Darstellungsvarianten auf, die nach einem rechten Mausklick auf die Liste gewählt werden können:

- Sortierung (nach Name, Reihenfolge in der Arbeitsdatei, Typ)
  Voreinstellung: Sortierung nach Namen
- Von der Labelanzeige kann auf Namensanzeige umgeschaltet werden, was in der unten wiedergegebenen Dialogbox geschehen ist.

9.1.1.2 Variablentypen

Die Variablen in der Anwärterliste einer interaktiven Grafik sind als **metrisch** (), **kategorial** () oder **systemeigen** () gekennzeichnet, wobei sowohl nominal- als auch ordinalskalierte Variablen (im Sinne des im Dateneditor zugewiesenen Messniveaus) als „kategorial“ bezeichnet werden. Mit *rein lokaler* Bedeutung für das aktuelle Diagramm kann der Typ einer Variablen auch in ihrem Kontextmenü geändert werden. Der Variablentyp entscheidet darüber, wie die Variablenausprägungen auf die Diagrammachsen abgebildet werden:

- Bei metrischen Variablen sind Intervalle zwischen zwei Werten bedeutsam.
- Bei kategorialen Variablen werden die Kategorien von Teilstrichen mit gleichmäßigem Abstand repräsentiert.
- Mit der systemeigenen Variablen **$case** kann z.B. ausgedrückt werden, dass jeder Fall der Arbeitsdatei auf der X-Achse dargestellt werden soll.

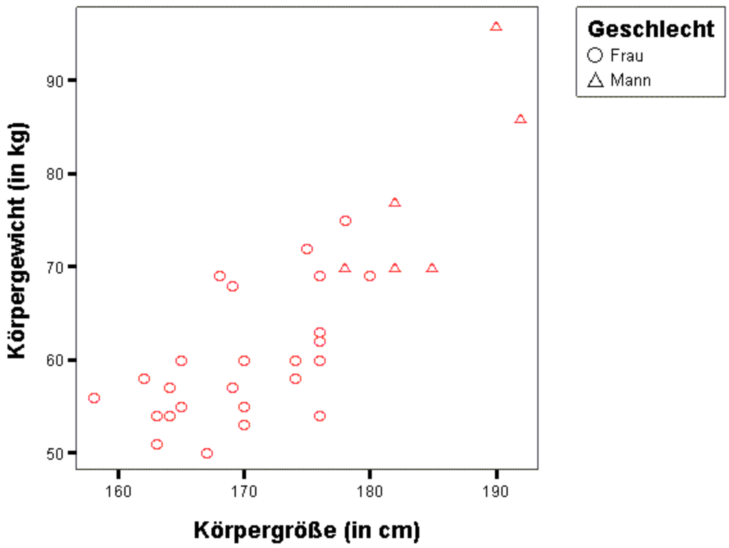
9.1.1.3 Zuordnung der Variablen

Die Variablen können per „Mauskralle“ (Ziehen und Ablegen) an ihren Bestimmungsort verfrachtet werden, was in folgender Dialogbox mit den Variablen GROESSE und GEWICHT geschehen ist:



9.1.1.4 Legendenvariablen

Im Streudiagramm kann durch **Farbe**, **Muster** oder **Größe** der Markierung eines Falles sein Wert bei **Legendenvariablen** dargestellt werden. Damit lässt sich z.B. der gemeinsame Effekt einer X-Achsen- und einer Legendenvariablen auf die Y-Achsen-Variable veranschaulichen.
In obiger Dialogbox wird eine Unterscheidung der beiden Geschlechts-Teilstichproben durch unterschiedliche Muster angefordert, was zu folgendem Ergebnis führt:



Man erkennt in der Graphik zumindest ansatzweise einen Geschlechtsunterschied hinsichtlich der Regressions*steigung*, der durch Unterschiede im Körperbau zu erklären ist:

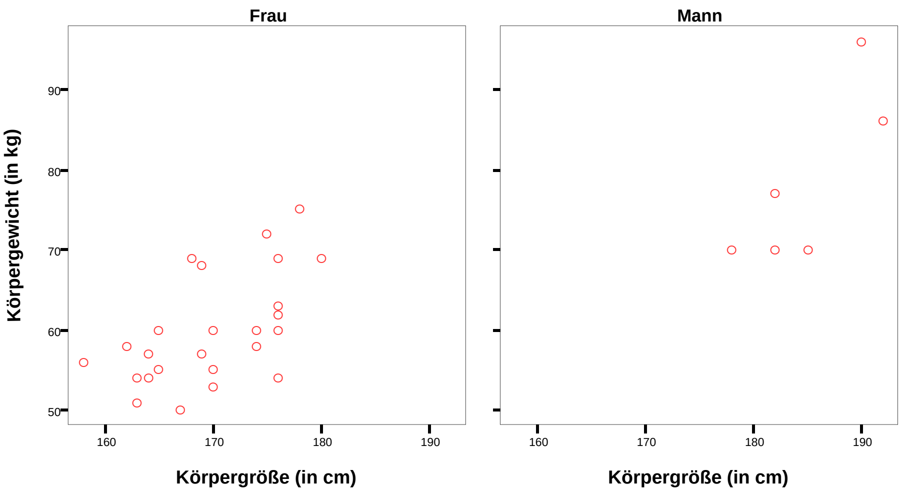


Bei zwei Männern mit 10 cm Größenunterschied ist ein stärkerer Gewichtsunterschied zu erwarten als bei zwei Frauen mit der selben Größendifferenz. Es ist also zu vermuten, dass Geschlecht den Effekt der Größe auf das Gewicht moderiert.

In der Regel wird man, wie in unserem Beispiel, kategoriale Legendenvariablen verwenden. Es sind jedoch auch *metrischen* Legendenvariablen möglich, z.B. mit Einfluss auf die Farbe oder Größe der Markierungen.

9.1.1.5 Feldvariablen

Eine Differenzierung verschiedener Teilstichproben kann auch mit Hilfe einer sogenannten **Feldvariablen** vorgenommen werden, die von kategorialem Typ sein muss. Für ihre Kategorien wird dabei jeweils ein eigenes Diagramm erzeugt, was in unserem Beispiel mit der Feldvariablen Geschlecht zu folgendem Ergebnis führt:

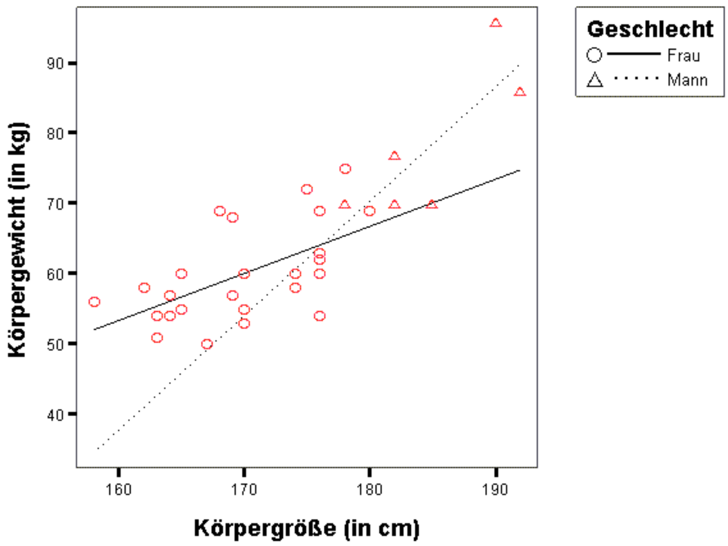


Bei Verwendung mehrerer Feldvariablen erhält man für jede Wertekombination ein eigenes Diagramm.

9.1.2 Anpassung einer Regressionsfunktion

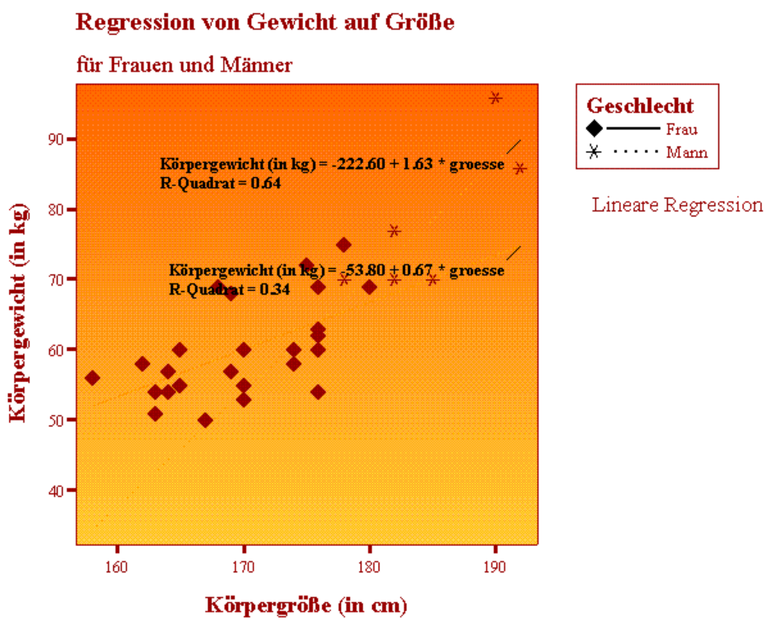
Auf dem Registerblatt **Anpassung** der Dialogbox zur Erstellung eines Streudiagramms kann man u.a. eine Regressionsfunktion anfordern, die global und/oder gruppenspezifisch an die Daten angepasst wird. Ferner sind Prognoseintervalle für Mittel- und Einzelwerte möglich.
Während der klassische SPSS-Grafikeditor, dessen Streudiagramm über den Menübefehl **Grafiken > Streudiagramm** nach wie vor verfügbar ist, auch eine *quadratische* Anpassungsfunktion anbietet, fehlt diese leider im interaktiven Grafikmodul.

Die folgende Abbildung zeigt, dass in Übereinstimmung mit der obigen Moderator-Spekulation die Regression von Gewicht auf Größe tatsächlich bei Männern steiler ausfällt als bei Frauen:



9.1.3 Titel und Optionen

Auf dem Registerblatt **Titel** können Beschriftungen vereinbart werden. Das Registerblatt **Optionen** bietet Diagrammvorlagen zur ästhetischen Gestaltung und ermöglicht das Strecken oder Stauchen der einzelnen Achsen. Mit Hilfe der Diagrammvorlage **Dante** entstand folgende Abbildung, deren Schönheitsfehler später noch repariert werden sollen:



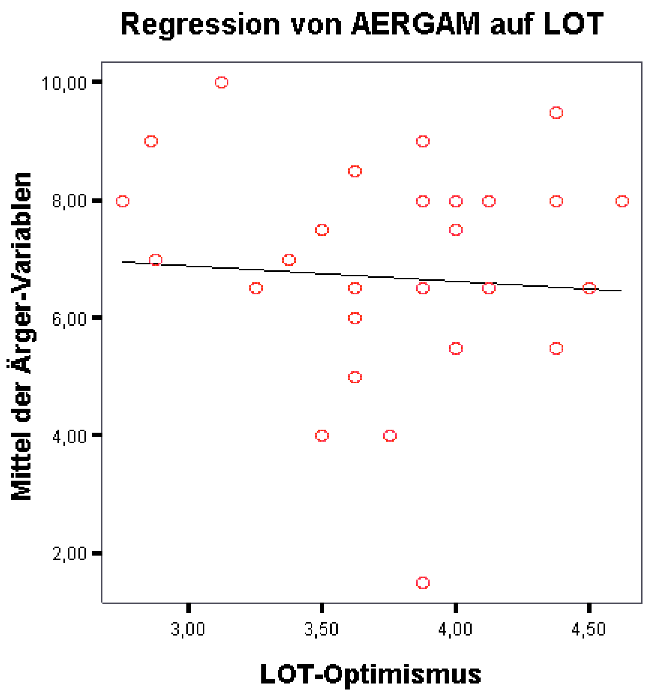
9.1.4 Weitere Hinweise

Um den gemeinsamen Effekt der beiden Regressoren Größe und Geschlecht auf die abhängige Variable Gewicht darzustellen, könnte man in der Dialogbox **Streudiagramm erstellen** auch ein *dreidimensionales* Streudiagramm anfordern. Allerdings ist im Falle des dichotomen Prädiktors Geschlecht die Gruppenaufteilung nach obigen Methoden weit übersichtlicher.

Die in der Dialogbox zur Erstellung einer interaktiven Grafik vorgenommenen Einstellungen können selbstverständlich später im Viewer-Fenster beliebig geändert werden.

9.1.5 Übung

Um Fehlentscheidungen aufgrund von technischen Fehlern zu vermeiden, sollten wir uns zu jedem statistischen Test die zugrundeliegenden deskriptiven Datenverhältnisse möglichst genau ansehen. Dies muss für die „gescheiterte" differentialpsychologische Hypothese (siehe Abschnitt 8) noch nachgeholt werden. Erzeugen Sie bitte dazu ein Streudiagramm mit den Variablen AERGAM und LOT.
Mit eingezeichneter Regressionsfunktion sollte es ungefähr so aussehen:



9.2 Interaktive Diagramme im Viewer-Fenster erstellen

Das in der letzten Übung zu erstellende Streudiagramm mit AERGAM und LOT kann bei einiger Phantasie als Überlagerung von zwei Verteilungen mit unterschiedlichen Regressionssteigungen gedeutet werden. Möglicherweise betätigt sich das Geschlecht wieder einmal als Moderator und beeinflusst den Zusammenhang von Optimismus und Ärger.

Während wir in Abschnitt 9.1 einen Moderatoreffekt über eine 2D-Grafik mit einer Legendenvariablen veranschaulicht haben, wollen wir nun das Phänomen mit einer 3D-Grafik darstellen. Dabei soll sich zeigen, ob der oft gegen 3D-Grafiken erhobene Vorwurf mangelnder Präzision und Klarheit berechtigt ist.

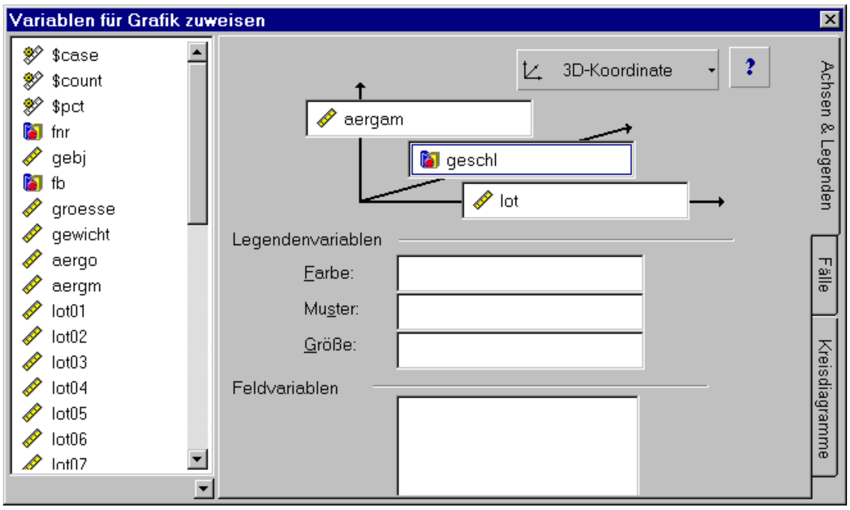
Wechseln Sie bitte nötigenfalls zum SPSS-Viewer, und fügen Sie eine leere 3D-Grafik ein mit

Einfügen > Interaktive 3-D Grafik

Aktivieren Sie über

Bearbeiten > Variablen zuweisen

bzw. mit dem Symbol  (linke obere Ecke der Bearbeitungszone) die Dialogbox **Variablen für Grafik zuweisen**. Bringen Sie auf der voreingestellten Registerkarte **Achsen & Legenden** die Variable LOT in die X-Achsen-Position, die Variable AERGAM in die Y-Achsen-Position und die Variable GESCHL in die senkrecht zur Zeichenebene liegende Achsenposition:



Schließen Sie anschließend die Dialogbox, um Platz zu sparen. Solange die aktuelle Arbeitsdatei weder verändert noch geschlossen wird, kann die Dialogbox wieder geöffnet werden, um die Rollen der Variablen neu zu verteilen.
Sie haben sicher schon festgestellt, dass Ihre Eintragungen sofort in die entstehende Zeichnung übernommen werden, die im Moment zwar beschriftete Achsen enthält aber noch keine graphischen Darstellungen. Beheben Sie bitte diesen Mangel, indem Sie über den Menübefehl **Einfügen > Zusammenfassung** oder über das Symbol das Element **Balken** ergänzen. Sie erhalten für Frauen und Männer jeweils eine Serie von Balken, wobei für jeden aufgetretenen LOT-Wert die Balkenhöhe den mittleren AERGAM-Wert der zugehörigen Teilstichprobe ausdrückt. Jede Balkenserie zeigt damit die bedingte Regression von AERGAM auf LOT, ohne Vorannahmen über den Funktionstyp zu machen.

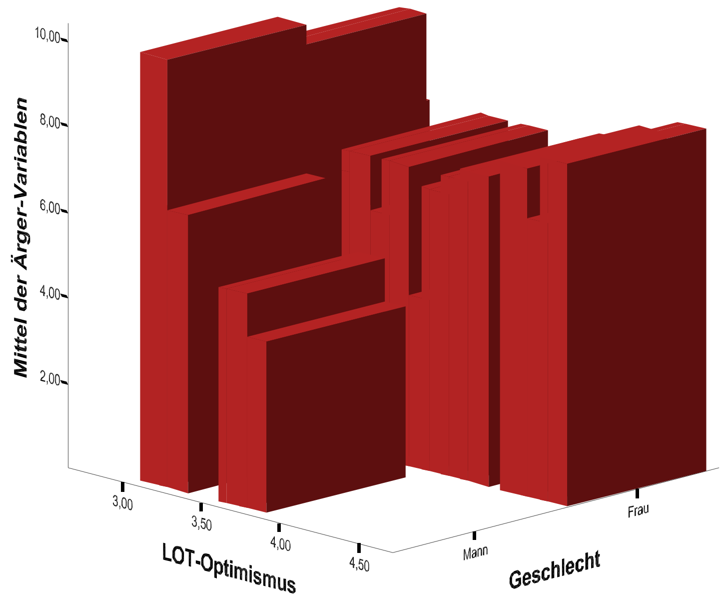
Als nächstes sollten Sie mit Hilfe der **3D-Palette** (über die Rädchen oder durch numerische Eingabe) einen möglichst günstigen Aussichtspunkt wählen (z.B. 345, 45).



Eine abgeschaltete 3D-Palette lässt sich mit **Ansicht > 3D-Palette** bzw. über das Symbol wieder aktivieren.
Die Symbole auf der 3D-Palette eröffnen folgende Bearbeitungsmöglichkeiten:

- Über die Achsen-Symbole mit den drehenden Pfeilen kann man zum vertikalen bzw. horizontalen Standardwert zurückkehren.
- Nach einem Klick auf die Faust mit dem aufgemalten Pfeil kann die Grafik auch per Maus rotiert werden.
- Wer will, kann nach einem Mausklick auf die Lampe auch die Lichtverhältnisse variieren.

Wenngleich die 3D-Darstellung deutlich ungenauer über die deskriptiven Verhältnisse informiert als die in Abschnitt 9.1 verwendete 2D-Grafik mit Geschlecht als Legendenvariable, wird doch der vermutete Moderatoreffekt sichtbar:



Während bei den Frauen offenbar kein Zusammenhang zwischen LOT und AERGAM besteht, zeigt sich bei den Männern ein Effekt im Sinne unserer differentialpsychologischen Hypothese. Allerdings sollten wir die Beobachtung sehr zurückhaltend interpretieren, weil unsere Stichprobe lediglich sechs Männer enthält.
Immerhin resultiert bei einer regressionsanalytischen Auswertung für den Moderatoreffekt[1] eine relativ kleine Überschreitungswahrscheinlichkeit (0,01):

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-19,356	11,285		-1,715	,098
	GESCHL * LOT	-7,883	2,860	-5,633	-2,756	,010
	Geschlecht	26,543	10,211	5,426	2,600	,015
	LOT-Optimismus	7,818	3,121	1,863	2,505	,019

a. Abhängige Variable: Mittel der Ärger-Variablen

Hier haben wir es aber **nicht** mit dem signifikanten Ergebnis eines statistischen Tests zu tun, sondern mit einem deskriptiven Maß zu einer interessanten Vermutung, die sich bei der explorativen Datenanalyse ergeben hat. Eine Testentscheidung über die Moderator-Hypothese ist nur in einer unabhängigen Stichprobe möglich.

9.3 Interaktive, abgetrennte und statische Diagramme

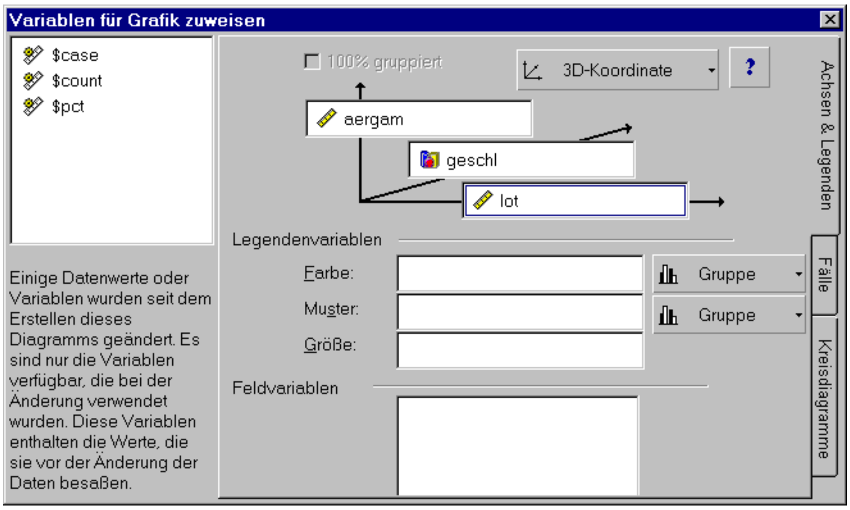
Die eben erstellte Grafik bleibt **interaktiv**, u.a. erkennbar an der Anzeige in der Viewer-Statuszeile (bei markierter, aber nicht geöffneter Graphik), solange die zugrundeliegende Ar-

[1] Über die Analyse von Moderatoreffekten mit Hilfe der SPSS-Regressions-Prozedur informiert eine elektronische Publikation des Rechenzentrums, die Sie auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend folgendermaßen erreichen:
Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen >
Statistische Spezialthemen > Moderatoranalyse per multipler Regression mit SPSS

beitsdatei unverändert geöffnet bleibt. Sie verliert diesen Status z.B. dann, wenn man die Arbeitsdatei um eine neue Variable erweitert. Dann resultiert eine **statische Grafik**, die zwar (wie die Diagramme des alten Grafik-Moduls) eine Änderung von Attributen vorhandener Objekte zulässt, aber z.B. keine neue Zuordnung von Variablen.
Es gibt einen Weg, den Verlust der Interaktivität (z.B. beim unweigerlich irgendwann erforderlichen Schließen oder Verändern der Arbeitsdatei) zu verhindern. Dazu muss nach dem Menübefehl

Bearbeiten > Optionen

auf dem Registerblatt **Interaktiv** der **Optionen**-Dialogbox im Bereich **Daten mit Diagramm speichern** die Voreinstellung **Nur ausgewertete Daten speichern** durch die Alternative **Daten mit Diagramm speichern** ersetzt werden. Die nach dieser Änderung erstellten interaktiven Graphiken, erhalten beim Schließen oder Ändern der Arbeitsdatei den Status **interaktiv (von Daten getrennt)**, der auch in der Viewer-Statuszeile eingeblendet wird (bei markierter, aber nicht geöffneter Graphik). Sie sind dann zwar von den Variablen der ursprünglich zugrundeliegenden Arbeitsdatei getrennt, beinhalten aber eine Kopie der im Diagramm verwendeten Variablen, die nach wie vor interaktiv umgeordnet werden können, z.B.:



9.4 Ändern von interaktiven Diagrammen

Um ein Diagramm im SPSS-Viewer verändern zu können, müssen Sie zunächst per Doppelklick den zuständigen Editor aktivieren. Wir wollen die oben erstellte Balkengrafik mit dem Moderationseffekt weiter gestalten.

9.4.1 Allgemeine Bearbeitungsmerkmale

Von den verschiedenen Möglichkeiten, eine Dialogbox zur Veränderung bestimmter Eigenschaften eines Objektes zu erhalten, sind zwei besonders zu empfehlen:

- Klicken Sie mit der rechten Maustaste auf das Objekt, und wählen Sie aus dem **Kontextmenü**.
  Entscheidet man sich z.B. im Kontextmenü zu einem Balken für die Option **Dieses geschl auswählen**, sind anschließend alle Balken der zugehörigen GESCHL-Teilstichprobe markiert. Öffnet man nun das Kontextmenü zu einem markierten Balken und wählt **Eigenschaften**, kann man anschließend für alle Balken der markierten Teilstichprobe eine alternative Farbe aussuchen.

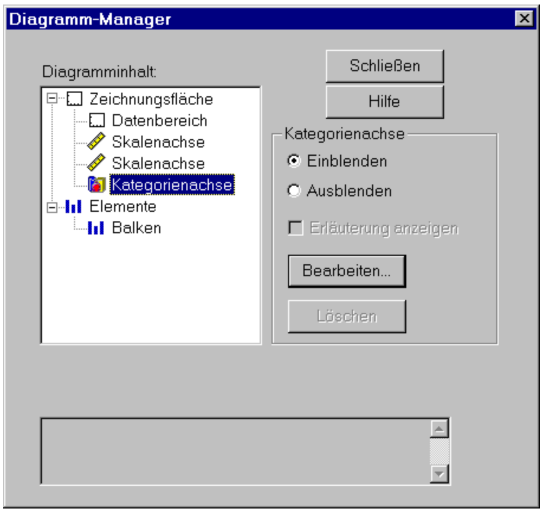
- Setzen Sie einen **Doppelklick** auf das Objekt.
  Z. B. erscheint nach einem Doppelklick auf einen Balken die Dialogbox **Balken**. Dort kann u.a. auf dem Registerblatt **Auswertungsfunktion** festgelegt werden, welchen Teilstichprobenkennwert die Balkenhöhe darstellen soll (z.B. Mittelwert, Median, Varianz).

Verschiedene Objekte (z.B. Überschriften, Legenden, Erläuterungen) können mit der Maus verschoben werden.

Die letzten Änderungen lassen sich über die Schalter rückgängig machen (mehrstufig) bzw. wieder herstellen.

9.4.2 Der Diagramm-Manager

Mit dem Menübefehl **Bearbeiten > Diagramm-Manager** bzw. über das Symbol starten Sie den Diagramm-Manager, der einen bequemen Zugriff auf viele Komponenten des Diagramms bietet:



Sie können z.B.

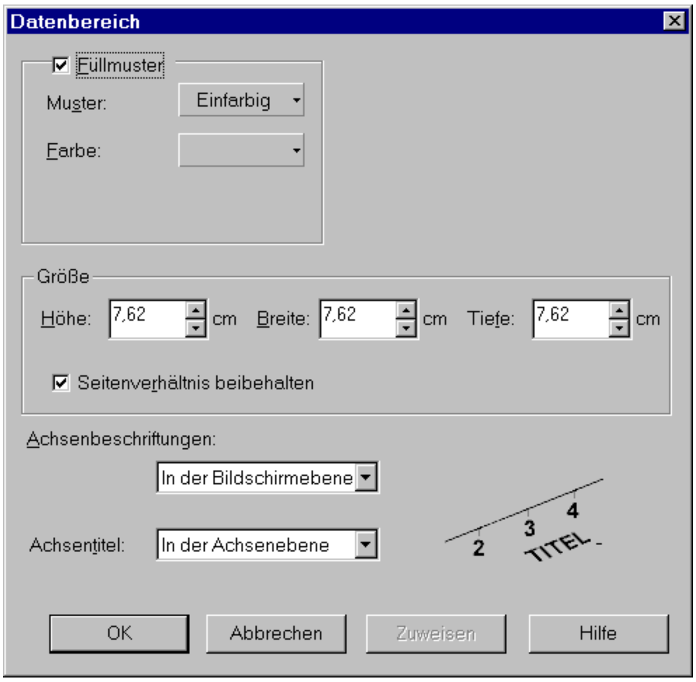
- die markierte Komponente aus- bzw. einblenden,
- die markierte Komponente bearbeiten oder löschen,
- die Anzeige der Erläuterung zu einem Element aus- bzw. einschalten.

Streichen Sie bitte mit Hilfe des Diagramm-Managers die Erläuterung zu den Balken („Balken zeigen Mittelwerte").

Im Diagramm-Manager finden Sie ggf. auch Warnungen zu Elementen, die Probleme machen oder z.B. nicht dargestellt werden können.

9.4.3 Der Datenbereich

Über den Menübefehl **Format > Datenbereich** bzw. über den Diagramm-Manager (**Datenbereich** markieren, **Bearbeiten** klicken) erreichen Sie eine Dialogbox zur Modifikation des Datenbereichs. Wählen Sie z.B. im Rahmen **Füllmuster** eine Wandtapete in dezentem Grau sowie **Achsenbeschriftungen** in der **Bildschirmebene** (bessere Lesbarkeit):



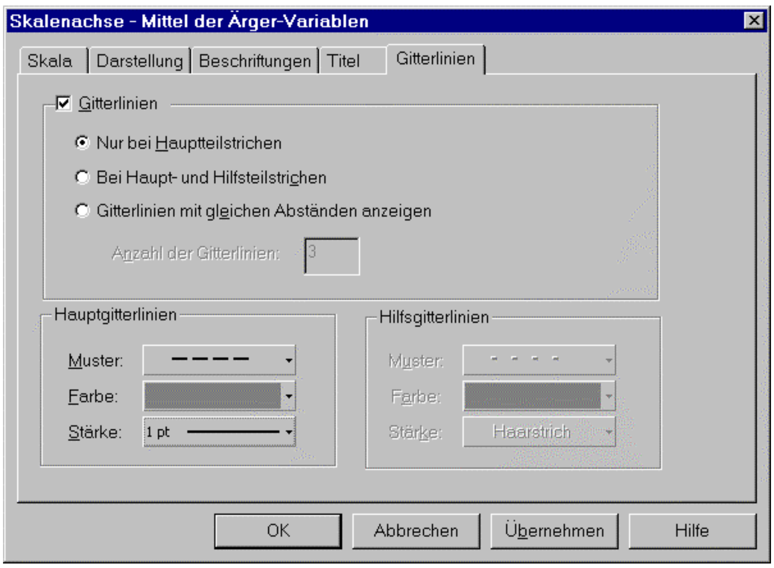
Sind Feldvariablen (s.o.) in Aktion, kann nach **Format > Datenbereich** in der Dialogbox **Felddatenbereich** die Anordnung der einzelnen Diagramme beeinflusst werden.

9.4.4 Modifikation der Achsen

Über den Diagramm-Manager sind Dialogboxen zur Gestaltung von Skalenachsen (für metrische Variablen) und Kategorienachsen zugänglich. Sie können u.a.:

- auf der Registerkarte **Skala** den darzustellenden Bereich festlegen,
- die Teilstriche gestalten,
- Wertbeschriftungen formatieren,
- Achsentitel formatieren,
- Gitterlinien festlegen.

Reduzieren Sie in unserem Beispiel bei der LOT- und der AERGAM-Achse die Anzahl der Dezimalstellen, und wählen Sie für die AERGAM-Achse gräuliche Gitterlinien mit Stärke 1:



9.4.5 Legenden

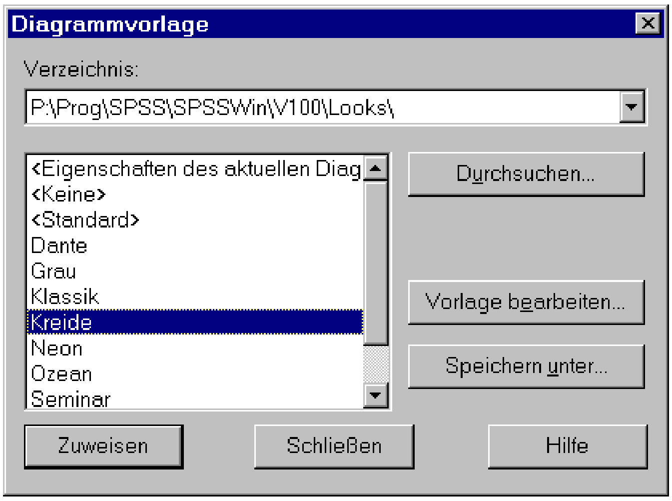
Während in früheren SPSS-Versionen die Diagramm-Legenden kaum modifiziert werden konnten, beschreibt das Handbuch zur interaktiven Grafik (SPSS 1999) nun auf 20 Seiten diverse Gestaltungsmöglichkeiten für Legenden zu kategorialen bzw. metrischen Variablen.

9.4.6 Diagrammvorlagen

Natürlich kann man Diagrammvorlagen nicht nur beim Erstellen einer interaktiven Grafik zuweisen (vgl. Abschnitt 9.1.3), sondern auch bei der Überarbeitung im Viewer. Dazu öffnet man über

Format > Diagrammvorlagen...

die folgende Dialogbox, die auch das **Bearbeiten** und **Speichern** von Vorlagen erlaubt:

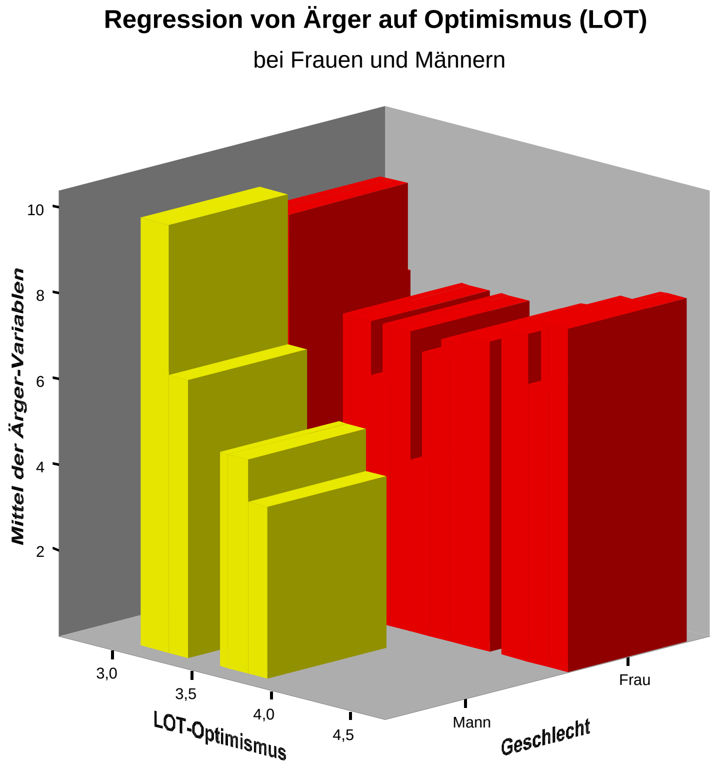


9.4.7 Weitere Gestaltungsmöglichkeiten

9.4.7.1 Beschriftungen

Mit Hilfe des **Einfügen**-Menüs lassen sich Titel, Untertitel und eine Erklärung ergänzen. Über das Text-Symbol kann man an beliebiger Stelle einen freien Text einfügen. Dieser lässt sich anschließend beliebig verschieben, was auch mit Titel, Untertitel und Erklärung klappt.
Das Ändern einer bestehenden Beschriftung wird mit einem Doppelklick eingeleitet. Einer markierten Beschriftung können nach **Format > Text** die üblichen Attribute zugeordnet werden.

Wenn wir in unserem 3D-Balkendiagramm noch Titel und Untertitel ergänzen, resultiert das folgende Ergebnis:

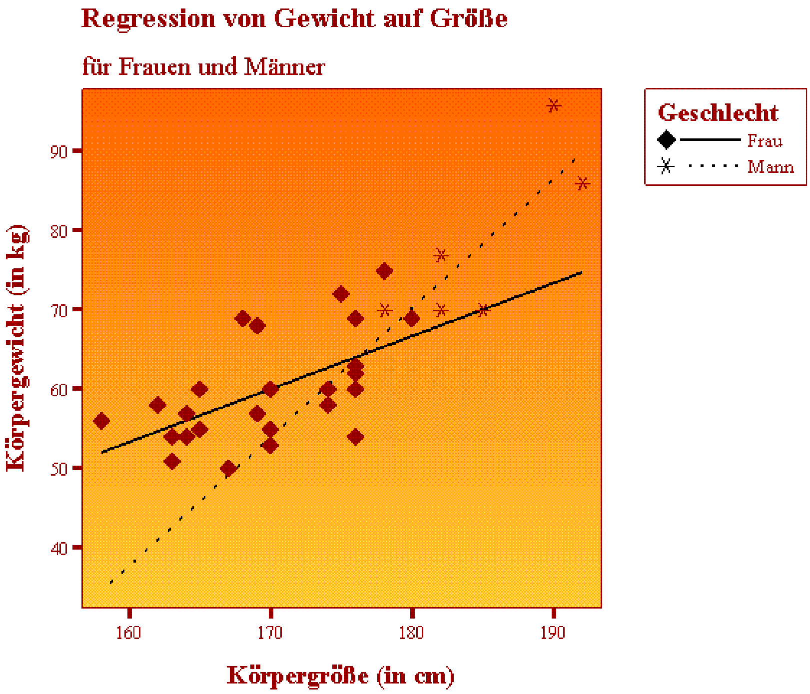


9.4.7.2 Farben, Muster, Linien, Markierungen etc.

In unserem Streudiagramm zur Regression von Gewicht auf Größe bei Frauen und Männern sind die beiden Regressionsgeraden aus farblichen Gründen nicht erkennbar. Außerdem sind die von SPSS in bester Absicht eingetragenen Regressionsgleichungen (mit Bezugslinien zur zugehörigen Regressionsgeraden) durch ihre Textfülle eher störend. Lösen sie diese beiden Probleme folgendermaßen:

- Öffnen Sie mit einem Rechtsklick das Kontextmenü zu den Regressionsgleichungen, und wählen Sie die Option **Beschriftung ausblenden**.
- Wählen Sie aus dem Kontextmenü zur weiblichen Regressionsgeraden den Eintrag **Optionen**, und legen Sie in der zugehörigen Dialogbox die Liniefarbe Schwarz und die Stärke Eins fest. Tun Sie das selbe auch für die männliche Regressionsgerade.

Wenn Sie im Diagramm-Manager noch die überflüssige Erläuterung zum Element **Regression** abschalten, sollten Sie in etwa folgendes Ergebnis sehen:



Nach dem Austauschen eines Symbols (Werkzeug) aktualisiert SPSS zwar die Legende, nicht aber den Datenbereich. Zwingen Sie das Programm in einem solchen Fall, seine Anzeige neu aufzubauen, z.B. durch zweimaliges Wechseln der Achsen mit und .

10 Fälle auswählen

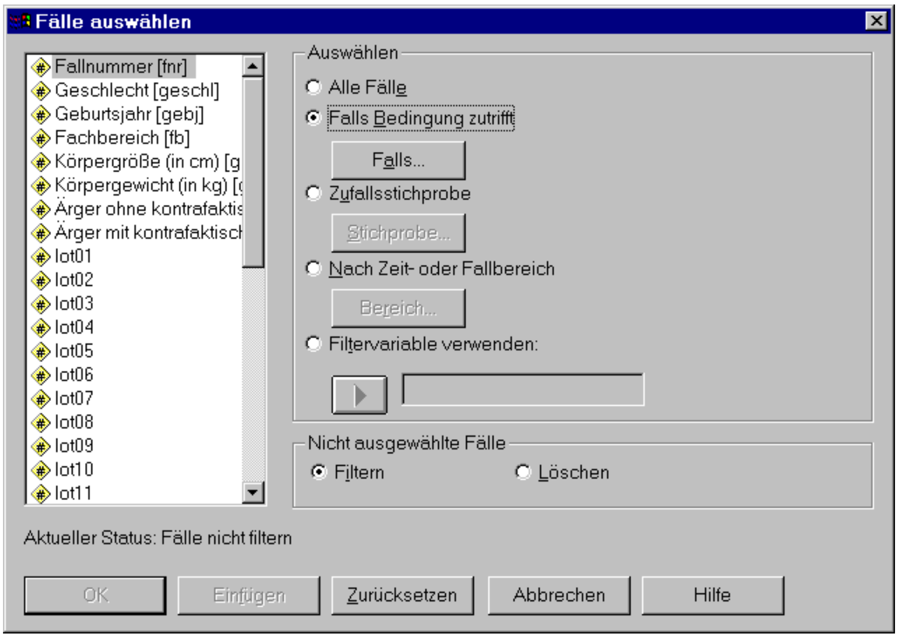
Es kommt durchaus vor, dass man sich bei einer statistischen Analyse auf eine Teilstichprobe beschränken möchte. Wir haben z.B. bei der graphischen Analyse der Regression von Gewicht auf Größe (siehe Abschnitt 9) festgestellt, dass man für Frauen und Männer jeweils eine eigene Regressionsgerade schätzen sollte. Dies lässt sich auf einfache Weise dadurch erreichen, dass bei der Regressionsrechnung jeweils eine Geschlechts-Teilstichprobe ausgeschlossen wird.

10.1 So wird's gemacht

SPSS erlaubt es, Fälle in Abhängigkeit von einer Bedingung temporär oder permanent aus der Arbeitsdatei auszuschließen. Die zuständige Dialogbox erreichen Sie über den Menübefehl:

Daten > Fälle auswählen...

Um eine Bedingung für die Teilnahme an den weiteren Auswertungen zu setzen, müssen Sie im Optionenfeld **Auswählen** die Alternative **Falls Bedingung zutrifft** markieren und anschließend die zugehörige Subdialogbox mit dem **Falls**-Schalter aktivieren:



Im **Falls**-Dialogfenster haben Sie die Möglichkeit, einen beliebigen logischen Ausdruck (vgl. Abschnitt 7.5.2) als Teilnahmekriterium zu definieren, z.B.:

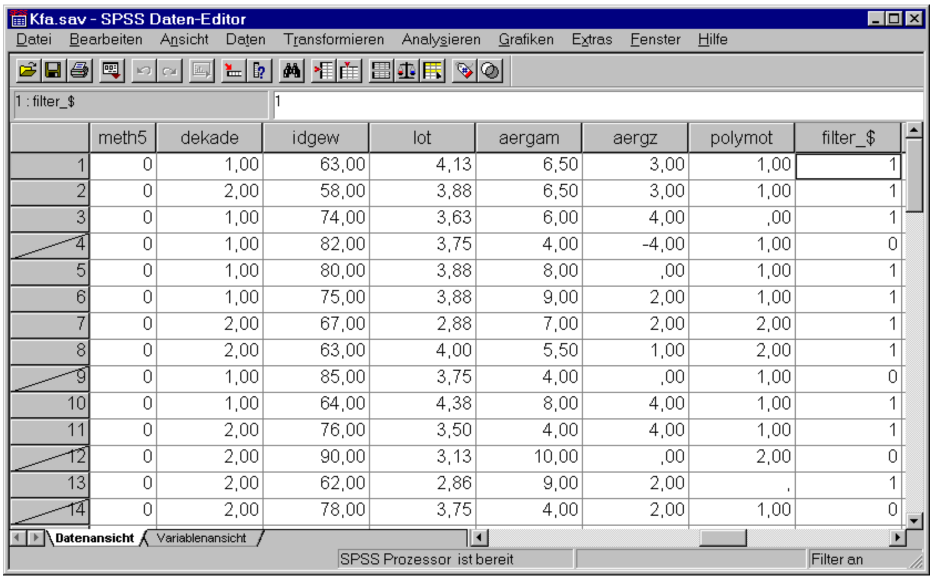


Wenn Sie nach erfolgreicher Definition des Teilnahmekriteriums **Weiter** machen, können Sie im Optionenfeld **Nicht ausgewählte Fälle** der Hauptdialogbox (siehe oben) entscheiden, was mit den Negativ-Fällen geschehen soll:

- **Filtern** SPSS erzeugt aufgrund Ihres logischen Ausdrucks eine Hilfsvariable namens FILTER_$ mit folgenden Werten:

  1 falls bei einem Fall der logische Ausdruck wahr ist,
  0 sonst (also auch bei unbestimmtem Ausdruck).

  Diese Variable wird als **Filter** aktiviert, d.h. bis zu einer Deaktivierung des Filters werden bei allen statistischen oder graphischen Analysen nur noch Fälle mit Wert 1 bei FILTER_$ einbezogen. Die in den einstweiligen Ruhezustand versetzten Null-Fälle sind im Datenfenster weiterhin vorhanden, ihre Fallnummern sind jedoch durchgestrichen:



	meth5	dekade	idgew	lot	aergam	aergz	polymot	filter_$
1	0	1,00	63,00	4,13	6,50	3,00	1,00	1
2	0	2,00	58,00	3,88	6,50	3,00	1,00	1
3	0	1,00	74,00	3,63	6,00	4,00	,00	1
~~4~~	0	1,00	82,00	3,75	4,00	-4,00	1,00	0
5	0	1,00	80,00	3,88	8,00	,00	1,00	1
6	0	1,00	75,00	3,88	9,00	2,00	1,00	1
7	0	2,00	67,00	2,88	7,00	2,00	2,00	1
8	0	2,00	63,00	4,00	5,50	1,00	2,00	1
~~9~~	0	1,00	85,00	3,75	4,00	,00	1,00	0
10	0	1,00	64,00	4,38	8,00	4,00	1,00	1
11	0	2,00	76,00	3,50	4,00	4,00	1,00	1
~~12~~	0	2,00	90,00	3,13	10,00	,00	2,00	0
13	0	2,00	62,00	2,86	9,00	2,00	.	1
~~14~~	0	2,00	78,00	3,75	4,00	2,00	1,00	0

  Wichtig: Filter wirken sich nur bei statistischen und graphischen Analysen aus. Bei Datentransformationen werden auch die ausgefilterten Fälle einbezogen. Wer eine bedingte Datentransformation benötigt, muss die Methoden aus Abschnitt 7.5 verwenden.

  Wenn ein Filter aktiv ist, wird dies in der Statuszeile angezeigt (siehe Abbildung).

  Um den Filter später zu deaktivieren, müssen Sie die Dialogbox **Fälle auswählen** erneut mit **Daten > Fälle auswählen** aufrufen und dann im **Auswählen**-Optionenfeld wieder den Ausgangszustand **Alle Fälle** markieren.

- **Löschen** Die Negativ-Fälle werden aus der Arbeitsdatei entfernt.
  Aus der *externen* Datei (z.B. auf der Festplatte) verschwinden die Fälle natürlich nicht. Wenn Sie allerdings das teilentleerte Datenfenster „sichern“, haben Sie eventuell anschließend ein kleines Problem.

Wird die obige Dialogbox **Fälle auswählen** mit **OK** quittiert, dann führt der SPSS-Prozessor im Hintergrund die folgenden Kommandos aus:

```
USE ALL.
COMPUTE filter_$=(geschl = 1).
VARIABLE LABEL filter_$ 'geschl = 1 (FILTER)'.
VALUE LABELS filter_$  0 'Nicht ausgewählt' 1 'Ausgewählt'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

Hinweise:

- Ist beim Sichern der Arbeitsdatei ein Filter aktiv, wird die zugrundeliegende Variable FILTER_$ mit abgespeichert. Beim nächsten Öffnen der Datei ist der Filter allerdings *nicht* aktiv, sondern muss nötigenfalls erneut vereinbart werden. Dazu muss aber lediglich die Variable FILTER_$ in der Dialogbox **Fälle auswählen** als **Filtervariable verwendet** werden. Weil es dabei auf den Namen der filternden Variable nicht ankommt, kann man beliebig viele Filter(variablen) in einer SPSS-Datendatei speichern.
  Außerdem kann man die einem Filter zugrundeliegende Syntax abspeichern und später wieder verwenden.
- Mit der oben dargestellten Dialogbox **Fälle auswählen** kann man auch eine *zufällige Teilstichprobe* ziehen oder eine Analyse auf die ersten *n* Fälle beschränken.

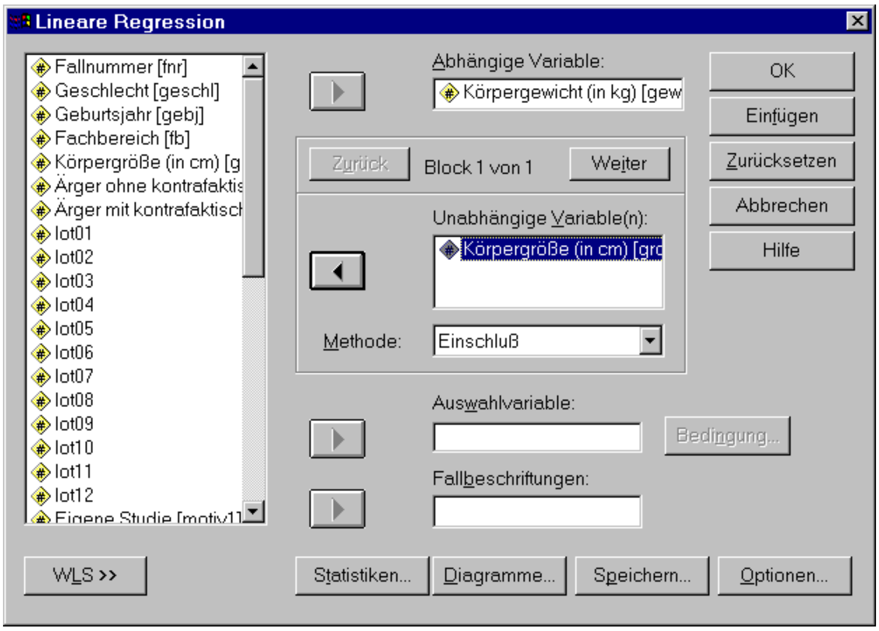
10.2 Übungen

1) Geschlechtsspezifische Regressionen von Gewicht auf Größe

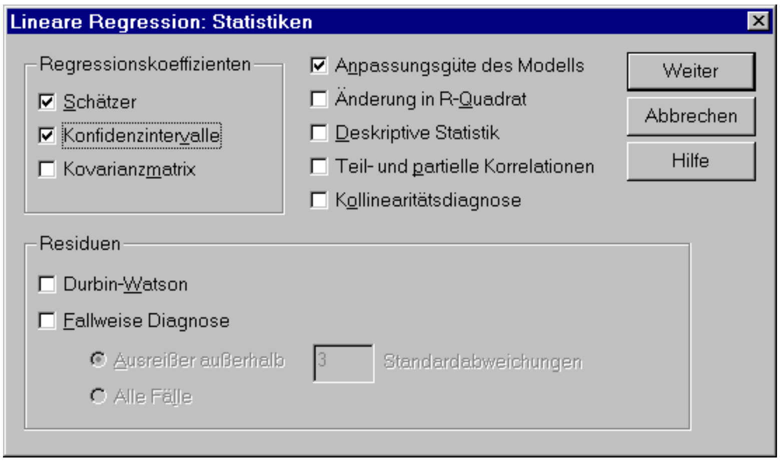
   Bestimmen Sie für Frauen und Männer getrennt die Regressionsgerade von Gewicht auf Größe. Die benötigte Regressions-Dialogbox erhalten Sie über:

   Statistik > Regression > Linear...

   Sie kann z.B. folgendermaßen ausgefüllt werden:



   Fordern Sie bitte in der **Statistiken**-Subdialogbox zusätzlich die Ausgabe von Konfidenzintervallen für die Regressionskoeffizienten an:



Weil die Teilstichproben nur sehr wenige Fälle enthalten, ist mit unerfreulich großen Konfidenzintervallen zu rechnen.
Wer sich ausführlich über die Regressionsanalyse mit SPSS informieren möchte, kann eine elektronische Publikation des Rechenzentrums zu diesem Thema auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend folgendermaßen erreichen:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen > Statistische Spezialthemen > Lineare Regressionsanalyse mit SPSS

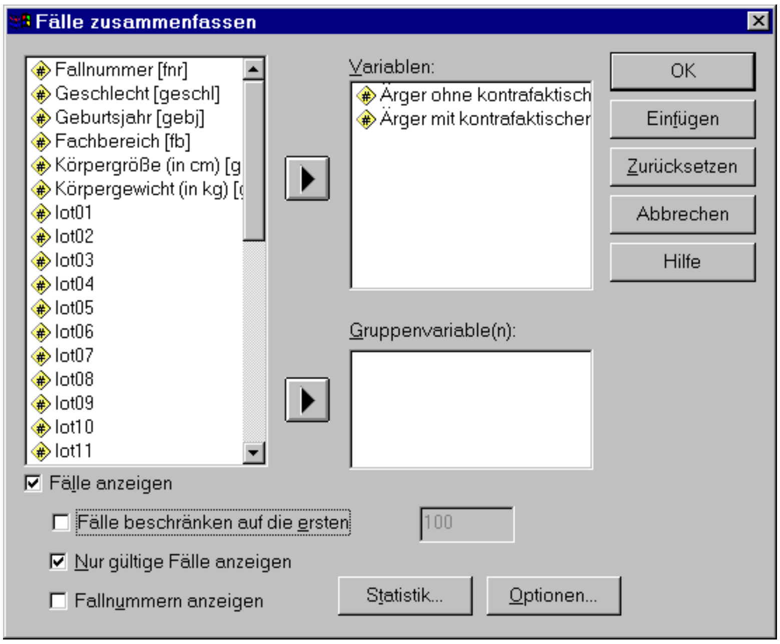
Hier wird u.a. diskutiert, wie man die statistischen Voraussetzungen einer linearen Regressionsanalyse (Linearität, Unabhängigkeit, Normalität und Homoskedastizität der Residuen) überprüfen kann.

2) Interessante Variablenausprägungen bei bestimmten Fälle auflisten

Gelegentlich benötigt man für eine bestimmte Teilmenge von Fällen eine übersichtliche Liste mit den Ausprägungen interessanter Variablen. Um z.B. für Personen mit negativem Ärgerzuwachs eine Liste mit den Variablen AERGO und AERGM zu erhalten, vereinbart man zunächst die Filterbedingung „AERGZ < 0“ und fordert dann über

Analysieren > Berichte > Fälle zusammenfassen...

die gewünschte Auflistung an:



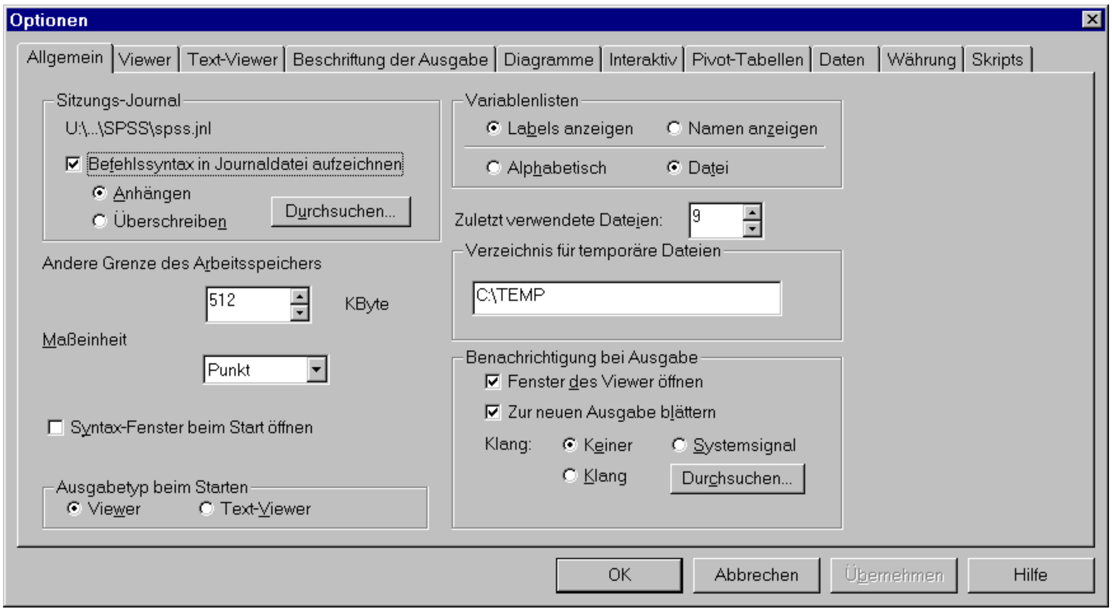
11 Einstellungen modifizieren

Das Standardverhalten von SPSS für Windows lässt sich auf vielfältige Weise den individuellen Bedürfnissen anpassen, was wir bei passender Gelegenheit auch schon getan haben.

Über den Menübefehl

Bearbeiten > Optionen...

erhalten Sie die folgende Dialogbox mit Optionen zur SPSS-Konfiguration:



Auf dem Registerblatt **Allgemein** sind folgende Optionen von Relevanz:

Sitzungs-Journal

Per Voreinstellung protokolliert SPSS alle Kommandos, die Sie während einer Sitzung per Dialogbox oder via Syntaxfenster abschicken, in einer sogenannten **Journaldatei**. Bei den Pool-PCs an der Universität Trier ist dies in der Regel die Datei:

U:\Eigene Dateien\SPSS\spss.jnl

Diese Journaldatei kann für Anwender(innen) mit „Mut zur SPSS-Syntax“ nach einem SPSS-Programmabsturz sehr nützlich sein, weil sie die Kommando-Äquivalente zu praktisch allen Arbeiten der verunglückten Sitzung enthält.

Per Voreinstellung wird beim Start einer SPSS-Sitzung eine vorhandene Journaldatei *nicht* überschrieben, sondern die neuen Kommandos werden am Ende angehängt. Falls die Datei zu groß wird, muss sie gelegentlich verkleinert oder gelöscht werden. Man kann aber auch im Rahmen **Sitzungs-Journal** der Karteikarte **Allgemein** den voreingestellten Öffnungsmodus **Anhängen** abändern auf **Überschreiben**. Dann wird die Journaldatei zu Beginn jeder Sitzung neu erstellt, wobei gegebenenfalls der alte Inhalt überschrieben wird.

Andere Grenze des Arbeitsspeichers

Während die meisten SPSS-Prozeduren den benötigten Hauptspeicher dynamisch anfordern, sind einige wenige auf eine manuelle Einstellung der **Anderen Grenze des Arbeitsspeichers** angewiesen. Die voreingestellten 512 KB brauchen Sie erst dann zu erhöhen, wenn SPSS explizit dazu auffordert. In diesem Fall ist die erhöhte Speicheranforderung sofort gültig.

Variablenlisten

Bei den Listen auswählbarer Variablen in Dialogboxen verwendet SPSS folgende Voreinstellungen:

- SPSS präsentiert die Variablen durch ihre Labels (falls vorhanden). Dabei werden die Variablenlisten aufgrund des begrenzten Platzangebotes oft recht unübersichtlich. Ein 50-stelliges Label, das auf ca. 20 Zeichen gekürzt werden musste, ist in der Regel weniger informativ als der vollständig sichtbare Variablenname (auf acht Stellen beschränkt). Mit der Option **Namen anzeigen** im Bereich **Variablenlisten** kann man auf die kompaktere Darstellung umschalten.
- Die Variablen sind angeordnet wie in der Arbeitsdatei, was in der Regel ein bequemes Arbeiten erlaubt. Gemeinsam zu analysierende und damit in Dialogboxen auszuwählende Variablen stehen nämlich oft in der Arbeitsdatei hintereinander. Bei der Arbeit mit einer unbekannten Datendatei findet man (namentlich bekannte) Variablen jedoch leichter bei alphanumerischer Sortierung. Im Rahmen **Variablenlisten** kann bei Bedarf das Sortierkriterium gewechselt werden.

Eine Änderung der Optionen für Variablenlisten wird leider erst beim nächsten Wechsel der Arbeitsdatei wirksam. Bei akutem Bedarf müssen Sie also die Arbeitsdatei schließen (**Datei > Neu > Daten**) und Ihre Datendatei dann neu öffnen.

12 Analyse von Kreuztabellen

Wir wollen die Hypothese prüfen, dass Frauen und Männer unterschiedliche Präferenzen bei der Wahl des Studienfachs haben.

Unsere Fachbereichs-Variable (FB) enthält Information über die Studienfächer der Untersuchungsteilnehmer(innen) auf einem angemessenen Aggregationsniveau. Ihre Werte stehen für die folgenden Fachbereiche der Universität Trier:

Fachbereich	Fächer
I	Pädagogik, Philosophie, Psychologie
II	Sprachorientierte Fächer
III	Historische und politische Wissenschaften
IV	BWL, Soziologie, VWL, Mathematik, Ethnologie
V	Jura
VI	Geowissenschaften

Nachdem die Begriffe aus der eingangs formulierten inhaltlichen Hypothese hinreichend präzisiert sind, können wir die empirisch zu prüfenden *Nullhypothese* formulieren:

Die Merkmale Geschlecht und Fachbereich sind unabhängig voneinander.

Die Unabhängigkeits-Behauptung der Nullhypothese bedeutet, dass sich aus dem Wissen über das Geschlecht eines Untersuchungsteilnehmers keinerlei Information über seine Fachbereichszugehörigkeit ableiten lässt, dass also die bedingten Fachbereichs-Verteilungen für beide Geschlechter identisch sind.

Zur Illustration des *Unabhängigkeits*begriffs wurde hier auf eine Verteilungs*homogenität* verwiesen. Später folgen noch einige Erläuterungen zu den beiden Begriffen und zu ihrer Beziehung.

Unsere Nullhypothesenformulierung ist „zweiseitig", wozu es auch gar keine Alternative gibt, weil die Fachbereichsvariable mehr als zwei Stufen hat. Bei (2×2)-Kreuztabellen sind auch einseitige Hypothesen möglich (siehe Abschnitt 12.3.3.2).

Da der Zusammenhang zwischen den beiden *nominalskalierten* Merkmalen Fachbereich und Geschlecht zu untersuchen ist, wählen wir als Auswertungsmethode die Kreuztabellenanalyse mit χ^2-Test.
Weil Kreuztabellenanalysen recht häufig benötigt werden, erläutert der vorliegende Abschnitt die wichtigsten statistischen Hintergründe und die Regeln für eine korrekte Interpretation der SPSS-Ergebnisse.

Leider erweist sich unsere KFA-Stichprobe bei näherer Betrachtung als ungeeignet zur Prüfung der Präferenz-Divergenz-Hypothese, denn

- Sie ist recht klein (geringe Teststärke).
- Die Stichprobe ist wenig repräsentativ, weil nur SPSS-Interessent(inn)en enthalten sind. Folglich sind manche Fachbereiche (z.B. III, V) fast nicht vertreten, weil dort SPSS kaum eingesetzt wird.

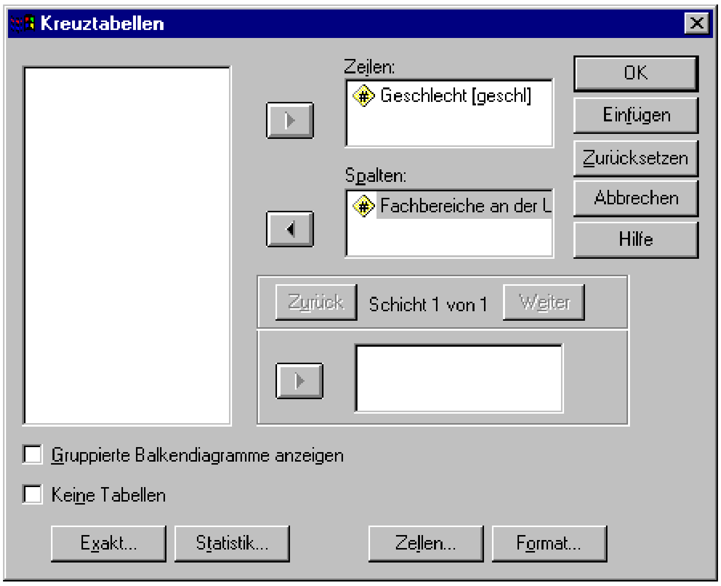
Daher wurde eine Zufallsstichprobe der Größe $n = 283$ aus der Kartei mit allen Studierenden der Universität Trier im WS 1993/94 gezogen[1]. Bei jeden Fall wurden die Variablen Geschlecht (GESCHL) und Fachbereich (FB) festgestellt.
Die SPSS-Datendatei **fbgeschl.sav** mit den beiden Variablen finden Sie an dem Ort, der im Vorwort für alle Kursdateien vereinbart wurde.

12.1 Beschreibung der bivariaten Häufigkeitsverteilung

Die SPSS-Dialogbox zur Analyse zweidimensionaler Kontingenztabellen erscheint nach dem Menübefehl:

Analysieren > Deskriptive Statistiken > Kreuztabellen...

Wir wählen GESCHL als Zeilen- und FB als Spaltenvariable:



In der **Zellen**-Subdialogbox kann man u.a. zeilen- und spaltenbezogene Prozentangaben für die Zellen der Kontingenztabelle anfordern:



Aufgrund dieser Spezifikationen erhalten wir für unsere Stichprobe die folgende Kreuztabelle[2]:

[1] Aufmerksame Leser(innen) werden zu Recht fragen, warum nicht *alle* Trierer Studierenden einbezogen wurden. Eine größere Stichprobe bringt stabilere Ergebnisse und hätte in dieser speziellen Situation kaum mehr „gekostet". Allerdings habe ich aus didaktischen Gründen eine Stichprobe mit „typischem" Umfang vorgezogen.

[2] Die Tabelle wurde mit dem Pivot-Editor durch Aufheben der Gruppierung Geschlecht etwas schlanker gemacht.

Geschlecht * Fachbereiche an der Universität Trier Kreuztabelle

		Fachbereiche an der Universität Trier						Gesamt
		I	II	III	IV	V	VI	
Frauen	Anzahl	29	26	18	22	26	23	144
	% von Geschlecht	20,1%	18,1%	12,5%	15,3%	18,1%	16,0%	100,0%
	% von FB	63,0%	66,7%	50,0%	31,0%	54,2%	53,5%	50,9%
Männer	Anzahl	17	13	18	49	22	20	139
	% von Geschlecht	12,2%	9,4%	12,9%	35,3%	15,8%	14,4%	100,0%
	% von FB	37,0%	33,3%	50,0%	69,0%	45,8%	46,5%	49,1%
Gesamt	Anzahl	46	39	36	71	48	43	283
	% von Geschlecht	16,3%	13,8%	12,7%	25,1%	17,0%	15,2%	100,0%
	% von FB	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Durch die Einträge in den Zellen wird die gemeinsame Verteilung der beiden Variablen GESCHL und FB beschrieben:

- Oben ... steht die absolute Häufigkeit der Zelle
  Z.B. befanden sich in der Stichprobe 17 Studenten aus dem Fachbereich I.
- In der Mitte ... steht der prozentuale Anteil der Zelle an allen Fällen in der zugehörigen Zeile.
  Z.B. gehörten von den 139 *männlichen* Untersuchungsteilnehmern 12,2% zum Fachbereich I.
  Diese auf die Zeile bezogenen relativen Häufigkeiten beschreiben also die bedingte Verteilung der Spaltenvariablen (FB) für einen festen Wert der Zeilenvariablen (GESCHL). Wir erhalten z.B. für die Männer die folgende bedingte Verteilung der Fachbereichs-Variablen:

I	II	III	IV	V	VI
12,2%	9,4%	12,9%	35,3%	15,8%	14,4%

- Unten ... steht der prozentuale Anteil der Zelle an allen Fällen in der zugehörigen Spalte
  Z.B. waren von den 46 Untersuchungsteilnehmern *aus dem Fachbereich I* 37% Männer.
  Diese auf die Spalte bezogenen relativen Häufigkeiten beschreiben also die bedingte Verteilung der Zeilenvariablen (GESCHL) für einen festen Wert der Spaltenvariablen (FB). Wir erhalten z.B. für den Fachbereich I die folgende bedingte Geschlechtsverteilung:

Frauen	63%
Männer	37%

In der **Zellen**-Subdialogbox können auch noch weitere Informationen zu den Zellen angefordert werden (z.B. der prozentuale Anteil der Zelle an der Gesamtstichprobe).

Beim Vergleich der fachbereichs-bedingten Geschlechtsverteilungen zeigen sich erhebliche Unterschiede:

- In den Fachbereichen I und II dominieren die Frauen mit einem Anteil von 63 bzw. 66,7%.
- Im Fachbereich IV sind die Frauen mit einem Anteil von nur 31% in der Minderheit.
- In den übrigen Fachbereichen III, V und VI zeigt sich ein relativ ausgeglichenes Geschlechtsverhältnis.

12.2 Die Unabhängigkeits- bzw. Homogenitätshypothese

Bei Hypothesen zu einer zweidimensionalen Kreuztabelle, also zur gemeinsamen Verteilung von zwei kategorialen Merkmalen, kann man den selben Sachverhalt mit unterschiedlichen wahrscheinlichkeitstheoretischen Begriffen beschreiben. Dies soll an unserem Beispiel demonstriert werden, damit Sie die Äquivalenz verstehen und ausnutzen lernen. Es ist ja generell sinnvoll, einen Sachverhalt aus verschiedenen Blickrichtungen zu betrachten.

1. Formulierung: Unabhängigkeitshypothese

H_0: Die Merkmale Geschlecht und Fachbereich sind unabhängig, d.h. die Wahrscheinlichkeit für jedes Verbundereignis (z.B. Mann im Fachbereich V) ist gleich dem Produkt aus den Wahrscheinlichkeiten der Randereignisse (im Beispiel: Mann, Fachbereich V).

H_1: Die Merkmale Geschlecht und Fachbereich sind abhängig, d.h. die Wahrscheinlichkeit für mindestens ein Verbundereignis ist ungleich dem Produkt aus den Wahrscheinlichkeiten der Randereignisse.

2. Formulierung: Homogenitätshypothese

H_0: Der Frauenanteil ist in allen Fachbereichen gleich.

H_1: Die Frauenanteile in den Fachbereichen sind verschieden.

Man kann leicht zeigen (vgl. Hartung 1989, S. 412):

Perfekte Homogenität liegt genau dann vor, wenn die Merkmale Geschlecht und Fachbereich unabhängig sind.

12.3 Testverfahren

12.3.1 Asymptotische χ^2 - Tests

Die bekannteste Prüfgröße zur Testung der Unabhängigkeits- bzw. Homogenitätshypothese ist die folgende χ^2_P **- Statistik nach Pearson**:

$$\chi^2_P := \sum_{i=1}^{z}\sum_{j=1}^{s}\frac{(n_{ij}-m_{ij})^2}{m_{ij}}, \quad \text{mit } m_{ij} = \frac{n_{i.}\cdot n_{.j}}{n}$$

Darin bedeuten:

z, s	=	Anzahl der Zeilen bzw. Spalten
n_{ij}	=	beobachtete Häufigkeit in Zelle *ij*
m_{ij}	=	unter der Nullhypothese erwartete Häufigkeit in Zelle *ij*
$n_{i.}$	=	beobachtete Häufigkeit in Zeile *i*
$n_{.j}$	=	beobachtete Häufigkeit in Spalte *j*
n	=	Umfang der Gesamtstichprobe

Wir wollen kurz überlegen, wie die angegebene Formel zur Berechnung der erwarteten Häufigkeiten m_{ij} unter der Nullhypothese zustande kommt. Zunächst soll die Wahrscheinlichkeit p_{ij} der Zelle *ij* unter der H_0 bestimmt werden. Da es sich hier um ein Verbundereignis aus zwei *unabhängigen* (H_0!) Einzelereignissen handelt (Zeile *i* und Spalte *j)*, ergibt sich p_{ij} als Produkt der Wahrscheinlichkeiten $p_{i.}$ bzw. $p_{j.}$ für die beiden verknüpften Einzelereignisse. Die Einzelwahrscheinlichkeiten $p_{i.}$ und $p_{j.}$ sind allerdings nicht bekannt, sondern müssen durch die entsprechen-

den relativen Häufigkeiten in den Daten geschätzt werden[1]. Z.B. wird die Wahrscheinlichkeit $p_{i.}$ zur Zeile *i* geschätzt durch die relative Häufigkeit der Zeile *i* in der Stichprobe:

$$\hat{p}_{i.} := \frac{n_{i.}}{n}$$

Analog ergibt sich die geschätzte Wahrscheinlichkeit $p_{.j}$ der Spalte *j*:

$$\hat{p}_{.j} := \frac{n_{.j}}{n}$$

Damit gilt für die geschätzte Wahrscheinlichkeit der Zelle *ij*:

$$\hat{p}_{ij} = \hat{p}_{i.} \cdot \hat{p}_{.j} = \frac{n_{i.}}{n}\frac{n_{.j}}{n} = \frac{n_{i.} \cdot n_{.j}}{n^2}$$

Um eine erwartete *Häufigkeit* zu erhalten, müssen wir jetzt nur noch die geschätzte Wahrscheinlichkeit mit der Stichprobengröße multiplizieren:

$$m_{ij} = \hat{p}_{ij} \cdot n = \frac{n_{i.} \cdot n_{.j}}{n^2} \cdot n = \frac{n_{i.} \cdot n_{.j}}{n}$$

Pearsons χ^2_P-Statistik ist offenbar ein Maß dafür, wie gut oder wie schlecht die erhobenen Daten mit der Nullhypothese zu vereinbaren sind: Im Zähler werden die quadrierten Abweichungen der beobachteten Häufigkeiten von den Erwartungswerten unter der H_0 aufsummiert. Durch das Quadrieren werden größere Diskrepanzen besonders stark gewichtet. Jede quadrierte Abweichung wird außerdem *normiert*, indem sie durch ihren erwarteten Wert dividiert wird. Steht etwa dem erwarteten Wert 5 die Häufigkeit 15 gegenüber, so resultiert die quadrierte und normierte Diskrepanz 20:

$$\frac{(15-5)^2}{5} = 20$$

Die selbe Abweichung einer beobachteten Häufigkeit 2010 vom erwarteten Wert 2000 erbringt jedoch sinnvollerweise nur eine quadrierte und normierte Diskrepanz von 0,005:

$$\frac{(2010-2000)^2}{2000} = 0{,}005$$

Es gilt offenbar, wie wir das in Abschnitt 8.1 von einer Prüfstatistik gefordert haben: Je größer der χ^2_P-Wert, desto unplausibler ist es, dass in der Population die Nullhypothese gilt.

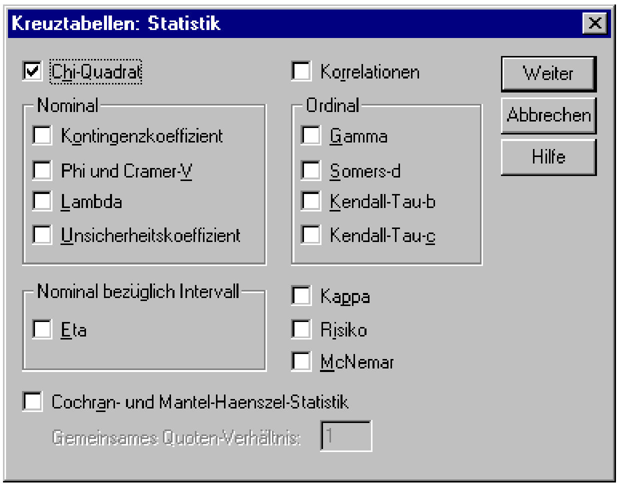
Außerdem erfüllt die χ^2_P-Teststatistik nach Pearson auch die Verteilungsbedingung aus Abschnitt 8.1, wenn auch nur approximativ. Unter der Nullhypothese ist die χ^2_P-Statistik asymptotisch, d.h. für $n \to \infty$, χ^2-verteilt mit df = (z - 1) · (s - 1) Freiheitsgraden.[2] Für unsere Kreuztabelle erhalten wir also z.B.: df = 1·5 = 5.

[1] Diese Formulierung geht davon aus, dass man *eine* Stichprobe gezogen und bei jedem Fall die *beiden* Merkmale Geschlecht und Fachbereich beobachtet hat. Ein anderes Stichprobenmodell läge vor, wenn man in jedem Fachbereich eine Stichprobe der festen Größe 50 gezogen und bei jedem Fall die *eine* Variable Geschlecht beobachtet hätte. Dann wären die Randwahrscheinlichkeiten der FB-Kategorien bekannt. Allerdings bleiben auch unter dem alternativen Stichprobenmodell alle vorgestellten Rechnungen und Entscheidungsregeln korrekt.

[2] In diesem Satz treten zwei Symbole mit ähnlicher Gestalt aber deutlich verschiedener Bedeutung auf: χ^2_P steht für eine (letztlich heuristisch definierte) Prüfgröße, mit χ^2 ist hingegen eine theoretische Verteilung gemeint.

Folglich kann mit Pearsons χ^2_P-Statistik nicht nur die Plausibilität der H_0 deskriptiv beurteilt werden, sondern es kann eine empirische Überschreitungswahrscheinlichkeit berechnet und nach den Regeln aus Abschnitt 8.1 ein Signifikanztest durchgeführt werden, d.h. man kann mit kontrollierter Irrtumswahrscheinlichkeit erster Art eine Entscheidung zwischen der H_0 und der H_1 treffen.

In SPSS wird der χ^2_P-Test mit dem Kontrollkästchen **Chi-Quadrat** in der **Kreuztabellen**-Subdialogbox **Statistik** angefordert:



Für unsere Daten erhalten wir folgendes Ergebnis:

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	18,191[a]	5	,003
Likelihood-Quotient	18,570	5	,002
Zusammenhang linear-mit-linear	3,197	1	,074
Anzahl der gültigen Fälle	283		

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 17,68.

Es ergibt sich ein χ^2_P-Wert von ca. 18,19, der bei df = 5 unter der H_0 eine Überschreitungswahrscheinlichkeit (**Asymptotische Signifikanz**) von ca. 0,003 hat, d.h. ein χ^2_P - Wert $\geq$ 18,19 bei df = 5 ist unter der H_0 extrem unwahrscheinlich. Insbesondere ist die empirisch ermittelte Überschreitungswahrscheinlichkeit deutlich kleiner als die üblicherweise akzeptierte Irrtumswahrscheinlichkeit von $\alpha = 0{,}05$. Folglich entscheidet sich der χ^2_P - Test klar für die H_1. Wir haben in Abschnitt 8.1 dieses Argumentationsmuster der Inferenzstatistik ausführlich erläutert.

Neben der χ^2_P-Statistik nach Pearson, die aus heuristischen Überlegungen hervorgegangen zu sein scheint, gibt es noch eine alternative Prüfgröße χ^2_{LQ}, die auf dem **Likelihood-Quotienten-Prinzip** basiert. Letztere ist unter der H_0 ebenfalls asymptotisch, d.h. für $n \to \infty$, χ^2 - verteilt mit df = (z-1)·(s-1) Freiheitsgraden, und trotz unterschiedlicher Herleitung sind beide Statistiken asymptotisch äquivalent, d.h. mit wachsender Stichprobengröße werden sie immer ähnlicher.

Während bei größeren Stichproben wegen der asymptotischen Äquivalenz die Entscheidung für eine der beiden Prüfgrößen beliebig ist, sprechen einige Befunde dafür, bei kleineren Stichproben die χ^2_P-Statistik nach Pearson wegen der besseren Verteilungsapproximation zu bevorzugen (siehe z.B. Hartung 1989, S. 439). Damit ist es also vertretbar, die χ^2_P-Statistik nach Pearson grundsätzlich gegenüber der Likelihood Ratio - Prüfgröße zu bevorzugen. SPSS liefert stets beide Prüfgrößen. In unserem Fall sind die Unterschiede geringfügig und für die Testentscheidung irrelevant.

Die Pearson- und die Likelihood-Quotienten-Statistik zur Beurteilung der Unabhängigkeits- bzw. Homogenitätshypothese sind nur **asymptotisch**, d.h. für $n \to \infty$, χ^2-verteilt. Für die Zulässigkeit der zugehörigen Hpothesentests setzt man üblicherweise voraus, dass alle **erwarteten** Häufigkeiten m_{ij} mindestens gleich **5** sind. SPSS protokolliert daher für jede Kreuztabelle die minimale erwartete Häufigkeit. In unserem Fall beträgt sie 17,682, so dass keine Einwände gegen Tests auf Basis der χ^2_P- bzw. χ^2_{LQ}-Statistik bestehen.

Manche Autoren formulieren etwas abgeschwächte Minimalforderungen an die erwarteten Häufigkeiten. Siegel (1976, S. 107) verlangt z.B. für χ^2_P-Tests mit df > 1, dass die beiden folgenden Bedingungen gelten müssen:

- Weniger als 20% der Zellen dürfen eine erwartete Häufigkeit kleiner als 5 haben.
- Keine Zelle darf eine erwartet Häufigkeit kleiner als 1 haben.

Neben den beiden Statistiken zur Prüfung der Unabhängigkeits- bzw. Homogenitätshypothese liefert SPSS unter der Bezeichnung **Zusammenhang linear-mit-linear** auch noch den χ^2_{MH}-Wert nach **Mantel-Haenszel** zur Beurteilung der **linearen** Beziehung zwischen den beiden Variablen. Diese Statistik darf nur interpretiert werden, *wenn beide Variablen Intervallskalenqualität besitzen*. Es handelt sich nämlich schlicht um die mit (n - 1) multiplizierte quadrierte Produkt-Moment-Korrelation zwischen den beiden Variablen:

$$\chi^2_{MH} := r^2(n-1)$$

Da wir zwei kategoriale Variablen betrachten, ist diese Statistik in unserem Fall also völlig sinnlos.

12.3.2 Exakte Tests

Für die (2 × 2)-Kreuztabellen gibt es seit Jahrzehnten mit dem **exakten Test von Fisher** eine glänzende Alternative zu den approximativen χ^2 – Tests. Wie sein Name sagt, kommt Fishers Test ohne Approximationen aus und ist daher bei jeder Stichprobe anwendbar. Erfreulicherweise bietet seit einigen Jahren das SPSS-System exakte Tests auch für beliebige (z × s)-Kreuztabellen.

Eine ausführliche Beschreibung der neuen statistischen Verfahren, die durch das SPSS-Zusatzmodul **Exact Tests** implementiert werden, finden Sie auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend über:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen > Statistische Spezialthemen > Exakte Tests mit SPSS

Allerdings sind die traditionellen asymptotischen Verfahren keinesfalls obsolet geworden, weil der exakte Test für (z × s)-Kreuztabellen wegen seines enormen Rechenaufwandes nur für kleine Stichproben durchführbar ist. Insgesamt steht für die meisten Situationen ein angemessenes Verfahren zur Verfügung:

- Wenn die Anwendbarkeitskriterien für die asymptotischen Verfahren erfüllt sind, sollten Sie den Pearson-Test verwenden.
- Anderenfalls sollten Sie einen exakten Test versuchen.

Wenn bei einer Kreuztabelle die Minimalanforderungen an die erwarteten Häufigkeiten *nicht* erfüllt sind, *und* der exakte Tests aufgrund des insgesamt zu großen Stichprobenumfangs nicht realisierbar, müssen Sie die verantwortlichen schwach besetzten Zeilen bzw. Spalten entweder löschen oder miteinander bzw. mit anderen Zeilen/Spalten zusammenlegen.

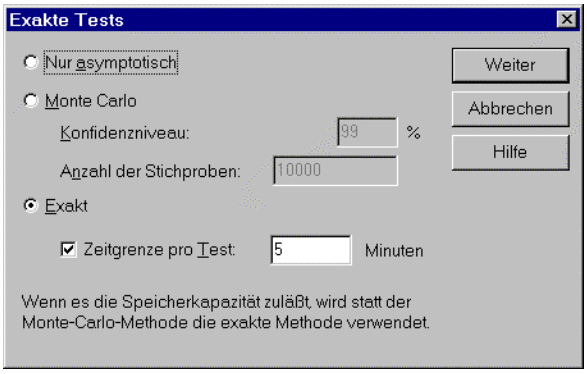
In einem Anwendungsbeispiel wollen wir die Daten aus dem ersten Abschnitt des SPSS-Handbuchs zum Modul **Exact Tests** (1996, S. 1) verwenden. Es handelt sich um Prüfungsergebnisse weißer, schwarzer, asiatischer und hispanoider Feuerwehrbewerber einer amerikanischen Kleinstadt.

Testergebnis * Hautfarbe Kreuztabelle

Anzahl

		Hautfarbe				
		Weiß	Schwarz	Asiatisch	Mittel- u. Südamerika	Gesamt
Testergebnis	Bestanden	5	2	2		9
	Unklar		1		1	2
	Durchgefallen		2	3	4	9
Gesamt		5	5	5	5	20

Wir wollen die Nullhypothese testen, dass die Prüfungsergebnisse nicht von der Hautfarbe abhängen.
Nach einem Mausklick auf den **Exakt**-Schalter in der Dialogbox zur Kreuztabellenanalyse können wir in der folgenden Subdialogbox die **exakte** Testmethode wählen:



Daraufhin erhalten wir neben den approximativen Ergebnissen auch exakte Überschreitungswahrscheinlichkeiten für die Pearson- und die Likelihood-Quotienten- – Prüfstatistik. Außerdem führt SPSS noch eine Verallgemeinerung des exakten Tests von Fisher durch, der in seiner klassischen Variante auf (2×2)-Tabellen beschränkt ist:

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)	Punkt-Wahrscheinlichkeit
Chi-Quadrat nach Pearson	11,556[a]	6	,073	,040		
Likelihood-Quotient	15,673	6	,016	,040		
Exakter Test nach Fisher	11,239			,040		
Zusammenhang linear-mit-linear	8,276[b]	1	,004	,004	,002	,001
Anzahl der gültigen Fälle	20					

a. 12 Zellen (100,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist ,50.

b. Die standardisierte Statistik ist 2,877.

Die approximativen χ^2 - Unabhängigkeitstests (Pearson und Likelihood-Quotient) sind nicht anwendbar, weil in allen 12 Zellen die erwartete Häufigkeit kleiner als 5 ist. Wer dieses Problem ignoriert, aber trotzdem weiß, dass der Pearson-Test dem Likelihood Ratio - Test im allgemeinen wegen der besseren Approximation vorzuziehen ist, gelangt zu einer falschen Testentscheidung, wie die Ergebnisse zur exakten Methode zeigen: Die korrekte Überschreitungswahrscheinlichkeit ist 0,04, was zur Ablehnung der Nullhypothese führt. Der asymptotische Pearson-χ^2-Test empfiehlt durch eine Überschreitungswahrscheinlichkeit von 0,07 hingegen, die Nullhypothese beizubehalten.

12.3.3 Besonderheiten bei (2×2)-Tabellen

12.3.3.1 Ein Klarer Fall für Fischers Test

Im beliebten Spezialfall der (2×2)-Tabelle ist Fishers Test nicht nur *exakt* für beliebige Stichproben, sondern er besitzt sogar unter allen „vernünftigen“, nämlich unter den sogenannten unverfälschten, Tests die besten Güteeigenschaften. Daher sollten Sie in dieser Situation grundsätzlich Fishers Test verwenden.

Die oben beschriebenen Rechenzeitprobleme bei exakten Tests für allgemeine $(z \times s)$-Kreuztabellen treten bei Fischers Test für die (2×2)-Tabelle *nicht* auf.

12.3.3.2 Einseitige Hypothesen

Bei einer (2×2)-Tabelle lässt sich im Unterschied zu allen anderen Tabellen die Unabhängigkeits- bzw. Homogenitätshypothese auch einseitig formulieren. Wenn wir uns z.B. beim Vergleich der Frauenanteile unter den Studierenden der Universität Trier auf die Fachbereiche III und IV beschränken, können wir die folgende einseitige Homogenitätshypothese aufstellen:

H_0: Der Frauenanteil ist im FB IV mindestens genauso groß wie im FB III.

H_1: Der Frauenanteil ist im FB IV kleiner als im FB III.

Aus den (z.B. per Filterbedingung, vgl. Abschnitt 10) eingeschränkten Beispieldaten erhalten wir folgende Ergebnisse:

Kreuztabelle

	Fachbereiche an der Universität Trier		
	III	IV	Gesamt
Frauen	18	22	40
	45,0%	55,0%	100,0%
	50,0%	31,0%	37,4%
Männer	18	49	67
	26,9%	73,1%	100,0%
	50,0%	69,0%	62,6%
Gesamt	36	71	107
	33,6%	66,4%	100,0%
	100,0%	100,0%	100,0%

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	3,689[b]	1	,055		
Kontinuitätskorrektur[a]	2,922	1	,087		
Likelihood-Quotient	3,643	1	,056		
Exakter Test nach Fisher				,061	,044
Zusammenhang linear-mit-linear	3,655	1	,056		
Anzahl der gültigen Fälle	107				

a. Wird nur für eine 2x2-Tabelle berechnet

b. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 13,46.

Wie wir bereits wissen, beträgt der Frauenanteil im FB III 50% und im FB IV 31%, die deskriptiven Statistiken fallen also klar im Sinne der Alternativhypothese aus. Der nach den obigen Überlegungen zu verwendende exakte Test von Fisher liefert für die *zweiseitige* Fragestellung eine Überschreitungswahrscheinlichkeit von 0,061, so dass die Nullhypothese beibehalten werden müsste. Bei *einseitiger* Testung erhalten wir jedoch eine Überschreitungswahrscheinlichkeit von 0,04, so dass die Nullhypothese verworfen werden kann.

Beachten Sie bitte abschließend noch, dass bei Fishers Test keinesfalls die einseitige Überschreitungswahrscheinlichkeit gerade die Hälfte der zweiseitigen Überschreitungswahrscheinlichkeit ist. Die in Abschnitt 8.1 für den Spezialfall des t-Tests angegebene Regel zur Berechnung der einseitigen Überschreitungswahrscheinlichkeit aus der zweiseitigen darf also nicht generalisiert werden.

12.3.3.3 Kontinuitätskorrektur nach Yates

Bei (2×2)-Tabellen berechnet SPSS traditionell auch eine χ^2_Y-Größe mit Kontinuitätskorrektur nach Yates. Sie soll bei kleineren Stichproben der Pearson-χ^2_P-Statistik überlegen sein. Gemäß Abschnitt 12.3.3.1 ist sie allerdings irrelevant für uns, weil in der (2×2)-Situation Fishers exakter Tests in jedem Fall vorzuziehen ist.

13 Fälle gewichten

Per Voreinstellung bezieht SPSS bei statistischen Auswertungen *alle* Fälle mit dem Gewicht *Eins* ein. In Abschnitt 10 haben Sie schon eine Möglichkeit kennen gelernt, Fälle aufgrund von Filterkriterien temporär oder permanent aus der Arbeitsdatei ausschließen. Nun erfahren Sie, wie man die Fälle individuell gewichtet, so dass sie bei statistischen Analysen unterschiedlichen Einfluss auf die Ergebnisse haben.

13.1 Beispiel

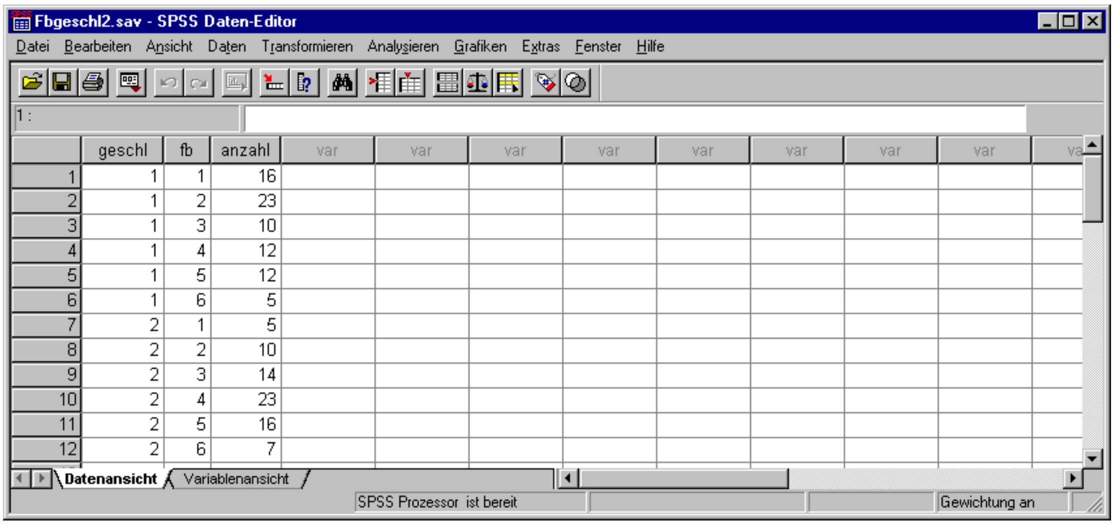
Die Möglichkeit, von 1 verschiedene Fallgewichte zu verwenden, d.h. z.B. einem Fall des Gewicht 16 zuzuschreiben und so zu tun, als seien 16 Fälle mit genau gleichen Variablenausprägungen in der Arbeitsdatei vorhanden, erscheint zunächst sinnlos. Aber erinnern wir uns an die (Geschlecht × Fachbereich) - Kreuztabelle aus Abschnitt 12. Zur Verwendung in einer späteren Übungsaufgabe betrachten wir hier eine strukturell identische Tabelle, die auf einer anderen Zufallsstichprobe der Größe $n = 153$ beruht:

	Fachbereich					
Geschlecht	I	II	III	IV	V	VI
Weiblich	16	23	10	12	12	5
männlich	5	10	14	23	16	7

Um mit den in Abschnitt 12 erklärten χ^2 - Tests anhand dieser Stichprobendaten prüfen zu können, ob in den Fachbereichen die Geschlechtsverteilungen verschieden sind, brauchen Sie nach unserem bisherigen Kenntnisstand eine Arbeitsdatei, in der z.B. 16 Fälle mit dem Geschlecht 1 und dem Fachbereich 1 enthalten sind, 23 Fälle mit Geschlecht 1 und Fachbereich 2 u.s.w. Wir haben jedoch lediglich die obige Tabelle zur Verfügung. Statt nun mühselig 153 Fälle im Dateneditor neu einzutippen, können wir von der Möglichkeit der Fallgewichtung folgendermaßen Gebrauch machen:

- Wir sorgen für ein leeres Datenfenster. Dort definieren wir die Variablen GESCHL (Geschlecht), FB (Fachbereich) und ANZAHL.
- Jede Zelle der (Geschlecht × Fachbereich) - Kreuztabelle wird im SPSS-Datenfenster als ein Fall behandelt.

  Der erste Fall erhält z.B. für die drei Variablen GESCHL, FB und ANZAHL die Werte 1, 1 und 16:



	geschl	fb	anzahl
1	1	1	16
2	1	2	23
3	1	3	10
4	1	4	12
5	1	5	12
6	1	6	5
7	2	1	5
8	2	2	10
9	2	3	14
10	2	4	23
11	2	5	16
12	2	6	7

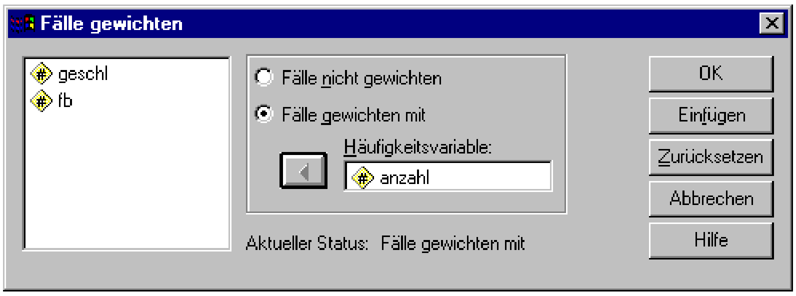
- Die Fälle werden mit der Variablen ANZAHL gewichtet. Damit tun wir z.B. so, als seien 16 Fälle mit dem Geschlecht 1 und dem Fachbereich 1 vorhanden gewesen. Aber das stimmt ja wirklich. Offenbar ist die Fallgewichtung doch nicht so sinnlos.

Um eine Gewichtsvariable zu vereinbaren rufen wir mit dem Menübefehl

Daten > Fälle gewichten...

eine sehr einfach zu bedienende Dialogbox auf. Sie besitzt ein Optionenfeld mit folgenden Alternativen:

- **Fälle nicht gewichten**
  Damit wird eine bestehende Gewichtung wieder aufgehoben.
- **Fälle gewichten mit**
  Die gewünschte Variable wird mit dem Transportschalter in die Position der **Häufigkeitsvariablen** gebracht, z.B.:



In der Dialogbox wird außerdem angezeigt, ob momentan eine Gewichtungsvariable vereinbart ist. Dieselbe Information erscheint auch in der Statuszeile des Datenfensters (siehe oben).

Beim Einsatz von Gewichtungsvariablen ist noch zu beachten:

- Zur Gewichtung kann natürlich nur eine numerische Variable verwendet werden, diese darf allerdings auch gebrochene Werte enthalten. Negative und fehlende Werte werden auf 0 gesetzt, d.h. die betroffenen Fälle werden nicht berücksichtigt, solange die Gewichtungsvariable aktiv ist.
- Ist beim Speichern der Arbeitsdatei eine Gewichtung aktiv, so wird diese mit abgespeichert und ist bei späterer Verwendung der Datendatei in Kraft.
- Bei der in diesem Abschnitt beschriebenen Anwendung der Gewichtungsoption wird dafür gesorgt, dass alle tatsächlich in der Studie vorhandenen Beobachtungen mit dem Gewicht 1 in die Kreuztabellenanalyse eingehen. Wenn die vorhandenen Beobachtungen individuelle Gewichte ($\neq 1$) erhalten, werden natürlich Signifikanztests erheblich beeinflusst. Auf jeden Fall muss dann die Gewichtungsvariable einen Mittelwert von 1 haben, d.h. die Summe der Gewichte muss gerade den Stichprobenumfang ergeben.

13.2 Übung

Prüfen Sie anhand der Daten aus der Tabelle am Anfang von Abschnitt 13.1 die Nullhypothese, dass die Merkmale Geschlecht und Fachbereich unabhängig sind.

14 Auswertung von Mehrfachwahl-Fragen

In Abschnitt 1.4.2.3 wurde betont, dass mit einer Mehrfachwahl-Frage nicht etwa *ein* mysteriöses Merkmal mit mehreren Ausprägungen erfasst wird, wie es wohl durch manche Köpfe bzw. Alpträume spukt, sondern schlicht *eine Familie* inhaltlich verwandter, aber logisch völlig unabhängiger Merkmale. Eine leichte Komplikation tritt erst auf, wenn zur Vereinfachung der Erfassung ein sparsames Set aus kategorialen Variablen definiert worden ist, das für viele Auswertungen erst „ausgepackt“ werden muss.

Grundsätzlich besteht kein Bedarf für spezielle Auswertungsverfahren für die mit Mehrfachwahlfragen erfassten Variablen. Es ist allerdings gelegentlich sinnvoll, eine Häufigkeits- oder Kreuztabellenanalyse für alle Mitglieder einer Familie inhaltlich verwandter Variablen (ob aus einer Mehrfachwahl-Frage entstanden oder wie auch immer) in gleicher Form auszuführen. Für diese Situation bietet SPSS gewisse Rationalisierungsmöglichkeiten, die in diesem Abschnitt vorgestellt werden sollen. Außerdem kann SPSS für die mit einem sparsamen Set aus kategorialen Variablen erfassten Merkmale Häufigkeits- und Kreuztabellenanalysen ohne vorheriges Auspacken durchführen.

In den Abschnitten 14.1 bis. 14.3 wird die Häufigkeits- bzw. Kreuztabellenanalyse für eine Familie von dichotomen Variablen beschrieben. In Abschnitt 14.4 wird demonstriert, wie man mit Hilfe einiger SPSS-Kommandos aus einem sparsamen Set kategorialer Variablen ein vollständiges Set dichotomer Variablen erzeugen kann.

14.1 Häufigkeitstabellen



Im Teil 4a unseres Fragebogens haben die Teilnehmer von fünf vorgegebenen Motiven, den SPSS-Kurs zu besuchen, die zutreffenden angekreuzt. Es liegt nahe, eine Übersicht zu erstellen, aus der für die einzelnen Motive hervorgeht, wie häufig sie gewählt worden sind. Natürlich können wir die Zustimmungs-Frequenzen bei den fünf Motiv-Variablen z.B. auch mit der längst bekannten Häufigkeitsanalyse (**Analysieren > Deskriptive Statistiken > Häufigkeiten...**) bestimmen lassen. SPSS bietet jedoch für solche *Gruppen zusammengehöriger Variablen* eine Prozedur an, welche die Zustimmungs-Häufigkeiten sowie einige zusätzliche Ergebnisse in besonders kompakter Form ausgibt. Wir erhalten für unsere Daten die folgende Tabelle:

```
Group $MOTIVE  Motive zur Kursteilname
     (Value tabulated = 1)

                                                           Pct of  Pct of
Dichotomy label                         Name       Count  Responses  Cases

Eigene Studie                           MOTIV1        23     56,1     76,7
Bewerbung um Stelle                     MOTIV2         1      2,4      3,3
Bewerbung um HIWI-Job                   MOTIV3         1      2,4      3,3
Interesse an der EDV                    MOTIV4         5     12,2     16,7
Interesse an Statistik                  MOTIV5        10     24,4     33,3
Keine Aussage zutreffend                KEINE          1      2,4      3,3
                                                 -------    -----    -----
                                Total responses       41    100,0    136,7

1 missing cases;  30 valid cases
```

Wir können etwa nachlesen, dass 23 Personen (= 76,7% aller validen Fälle) dem ersten Motiv zugestimmt haben. Diese 23 positiven Antworten machen 56,1% der insgesamt 41 von allen Teilnehmern geäußerten Zustimmungen aus. Ein Fall, auf den wir später noch eingehen müssen, fand keines der vorgegebenen Motive für sich passend. Seine „Zustimmung“ bei der Variablen KEINE stört die Interpretierbarkeit der Spalte `Pct of Responses`, die aber ohnehin weit weniger relevant ist als die benachbarte Spalte `Pct of Cases`.

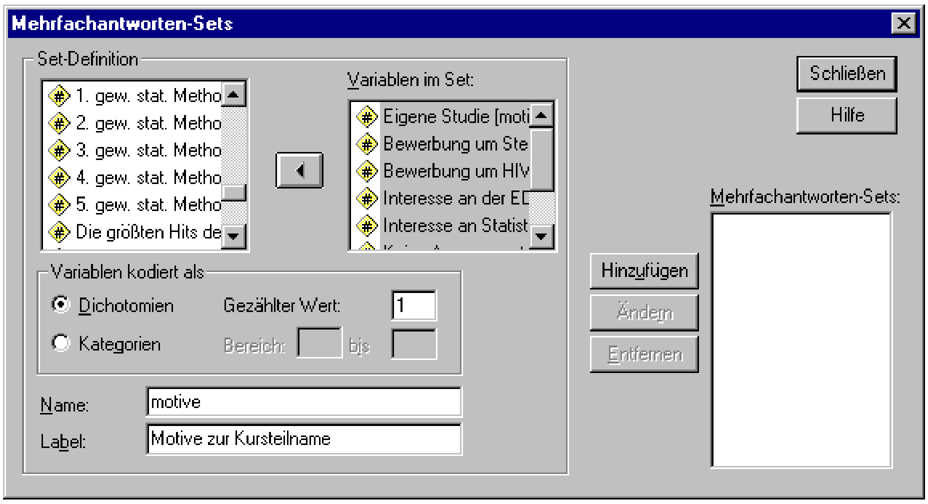
Wie die erste Zeile in obiger Ausgabe zeigt, wurde hier eine Prozedur für die **Variablengruppe** $MOTIVE ausgeführt, die natürlich zuvor definiert werden muss. Wählen Sie dazu den Menübefehl:

Analysieren > Mehrfachantworten > Sets definieren...

In der nun erscheinenden Dialogbox sind folgende Aktionen nötig:

- Befördern Sie die Variablen MOTIV1 bis MOTIV5 sowie KEINE in die Liste **Variablen im Set**.
- Tragen Sie im Rahmen **Variablen kodiert als** für die bei uns zutreffende **dichotome** Option die 1 als zu zählenden Wert ein.
- Vereinbaren Sie für das Set den Namen MOTIVE und das Label „Motive zur Kursteilname".

Danach müsste Ihre Dialogbox so aussehen:

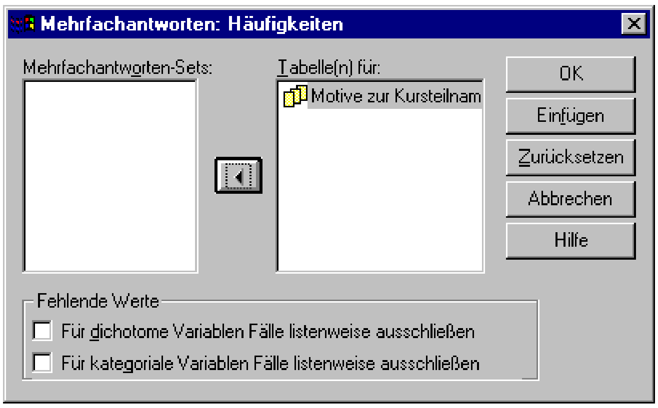


Nehmen Sie abschließend mit **Hinzufügen** die neue Gruppe in die Liste der Mehrfachantworten-Sets auf, und schließen Sie die Dialogbox.

Nun können Sie obige Ergebnisse leicht anfordern mit dem Menübefehl

Analysieren > Mehrfachantworten > Häufigkeiten...

und der zugehörigen Dialogbox:



Bei der Ausgabe handelt es sich um keine Pivot-Tabelle, sondern um eine Testausgabe alten Stils, für die im Viewer-Fenster ein eigener Editor zuständig ist. Diese Textausgaben alter Prozeduren, die noch keine Pivot-Tabellen beherrschen, sind von den Tabellen im Text-Viewer

zu unterschieden, die in der Regel durch das Umwandeln von Pivot-Tabellen entstanden sind (vgl. Abschnitt 8.6).
Wenn das SPSS-Zusatzmodul **Tables** zur Verfügung steht, können Häufigkeitstabellen für Mehrfachantworten-Sets auch im Pivot-Format erstellt werden (siehe Abschnitt 14.3).

! Lässt man die Variable KEINE zur Antwort „Für mich trifft keine der obigen Aussagen zu" bei der Set-Definition weg, resultiert die folgende Häufigkeitstabelle:

```
Group $MOTIVE  Motive zur Kursteilnahme
     (Value tabulated = 1)

                                                        Pct of  Pct of
Dichotomy label                         Name      Count  Responses  Cases

Eigene Studie                           MOTIV1        23     57,5    79,3
Bewerbung um Stelle                     MOTIV2         1      2,5     3,4
Bewerbung um HIWI-Job                   MOTIV3         1      2,5     3,4
Interesse an der EDV                    MOTIV4         5     12,5    17,2
Interesse an Statistik                  MOTIV5        10     25,0    34,5
                                                 -------    -----   -----
                                Total responses       40    100,0   137,9

2 missing cases;  29 valid cases
```

Bei dieser Tabelle erstaunt, dass nur **29** valide Fälle gemeldet werden, obwohl sich in unserer KFA-Datendatei **30** Fälle mit vollständig vorhandenen MOTIV-Werten befinden. Des Rätsels Lösung ist eine m.E. fragwürdige SPSS-Eigenart bei der Analyse von Mehrfachantworten-Sets aus dichotomen Variablen: Als „valide" gelten genau diejenigen Fälle, die bei mindestens einer Variablen den zu zählenden Wert besitzen (bei uns also die 1). Daher wird neben dem Fall 13 mit SYSMIS bei den Variablen MOTIV1 bis MOTIV5 auch der dritte Fall ausgeschlossen, der *alle Motive verneint*, also den Fragebogenteil 4a mit einem einzigen Kreuz bei „Für mich trifft keine der obigen Aussagen zu" korrekt bearbeitet hat. Wenn SPSS in obiger Ausgabe z.B. zum Motiv 1 meldet, dass 79,3% der Fälle (23 von 29) zugestimmt hätten, ist dies schlicht **falsch**.

Der Mangel in obiger Ausgabe wurde aufgrund der Meldung "`2 missing cases`" entdeckt. Sie sollten grundsätzlich bei allen SPSS-Ausgaben die protokollierten Fallzahlen überprüfen, weil sehr viele technische Fehler sich durch eine zu niedrige oder zu hohe Anzahl auswertbarer Fälle verraten. Im aktuellen Beispiel ist SPSS für den „Fehler" verantwortlich, in der Regel werden Sie auf diese Weise Ihre eigenen Fehler entdecken.

Die einzige Möglichkeit, definierte Mehrfachantworten-Sets zu speichern, besteht darin, die zur Häufigkeitsanalyse bzw. zur anschließend beschriebenen Kreuztabellenanalyse gehörige Syntax zu sichern. In den korrespondierenden SPSS-Kommandos sind die Set-Definitionen nämlich enthalten, z.B.:

```
MULT RESPONSE
  GROUPS=$motive 'Motive zur Kursteilnahme' (motiv1 motiv2 motiv3 motiv4
  motiv5 keine (1))
  /FREQUENCIES=$motive  .
```

14.2 Kreuztabellen

Wenn wir uns für Geschlechtsunterschiede bei der Zustimmung zu den einzelnen Motiven interessieren (z.B.: „Wer interessiert sich mehr für Statistik?"), sind genau *fünf* (2×2)-Tabellen zu analysieren. Über den aus Abschnitt 12 bekannten Menübefehl **Analysieren > Deskriptive Statistiken > Kreuztabellen** erhalten wir z.B. für das Statistik-Motiv (Nummer Fünf) folgendes Ergebnis:

Interesse an Statistik * Geschlecht Kreuztabelle

			Geschlecht		Gesamt
			Frau	Mann	
Interesse an Statistik	Nein	Anzahl	15	5	20
		% von Interesse an Statistik	75,0%	25,0%	100,0%
		% von Geschlecht	62,5%	83,3%	66,7%
	Ja	Anzahl	9	1	10
		% von Interesse an Statistik	90,0%	10,0%	100,0%
		% von Geschlecht	37,5%	16,7%	33,3%
Gesamt		Anzahl	24	6	30
		% von Interesse an Statistik	80,0%	20,0%	100,0%
		% von Geschlecht	100,0%	100,0%	100,0%

Weil die Motiv-Variablen nur zwei Ausprägungen haben, ist in dieser Tabelle die erste Zeile redundant. Es genügt zu wissen, dass 37,5% der Frauen und 16,7% der Männer ein Interesse an der Statistik angegeben haben. Durch Verzicht auf die redundanten Zeilen erhält man eine sehr kompakte Darstellung der fünf (2×2)-Tabellen zu Geschlechtsunterschieden bei den Kursmotiven:

```
                        GESCHL
                Count  |Frau     Mann
              Row pct  |                   Row
              Col pct  |                   Total
                       |     1  |     2  |
$MOTIVE         -------+--------+--------+
             MOTIV1    |    19  |     4  |    23
  Eigene Studie        |  82,6  |  17,4  |  76,7
                       |  79,2  |  66,7  |
                       +--------+--------+
             MOTIV2    |     1  |     0  |     1
  Bewerbung um Stelle  | 100,0  |    ,0  |   3,3
                       |   4,2  |    ,0  |
                       +--------+--------+
             MOTIV3    |     0  |     1  |     1
  Bewerbung um HIWI-Jo |    ,0  | 100,0  |   3,3
                       |    ,0  |  16,7  |
                       +--------+--------+
             MOTIV4    |     3  |     2  |     5
  Interesse an der EDV |  60,0  |  40,0  |  16,7
                       |  12,5  |  33,3  |
                       +--------+--------+
             MOTIV5    |     9  |     1  |    10
  Interesse an Statist |  90,0  |  10,0  |  33,3
                       |  37,5  |  16,7  |
                       +--------+--------+
             KEINE     |     1  |     0  |     1
  Keine Aussage zutref | 100,0  |    ,0  |   3,3
                       |   4,2  |    ,0  |
                       +--------+--------+
               Column       24        6       30
                Total     80,0     20,0    100,0

Percents and totals based on respondents
30 valid cases;  1 missing cases
```

Beachten Sie bitte: Dies ist **nicht eine** (5×2)-Kontingenztabelle, **sondern dies sind fünf** (2×2)-Kontingenztabellen. In der vorletzten Zeile befindet sich etwa die Essenz der MOTIV5 × GESCHL - Kontingenztabelle.

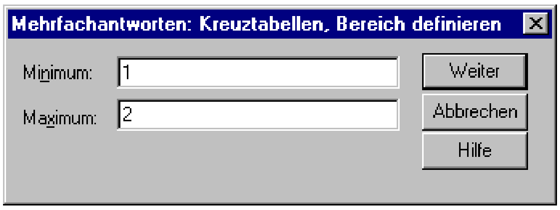
Sie können die Dialogbox zu obiger „Kombi"-Tabelle anfordern mit

Analysieren > Mehrfachantworten > Kreuztabellen...

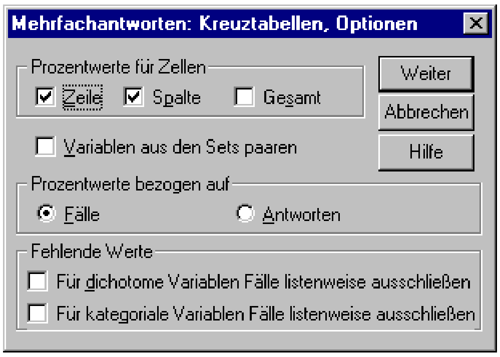
Befördern Sie das Mehrfachantworten-Set $MOTIVE in die **Zeilen**:



Für die **Spalten**-Variable GESCHL müssen Sie noch den folgenden **Bereich definieren**:



Die spalten- und zeilenrelativierten Prozentangaben werden in der **Optionen**-Subdialogbox angefordert:



Auch bei dieser Kontingenzanalyse ist die in Abschnitt 14.1 gerügte MD-Konzeption der SPSS-Mehrfachantworten-Auswertung zu beachten. Wäre nicht die Variable KEINE Mitglied im Mehrfachantworten-Set $MOTIVE, dann würde SPSS in der Kombi-Tabelle nur noch diejenigen Fälle berücksichtigen, die mindestens ein Motiv bejaht haben.

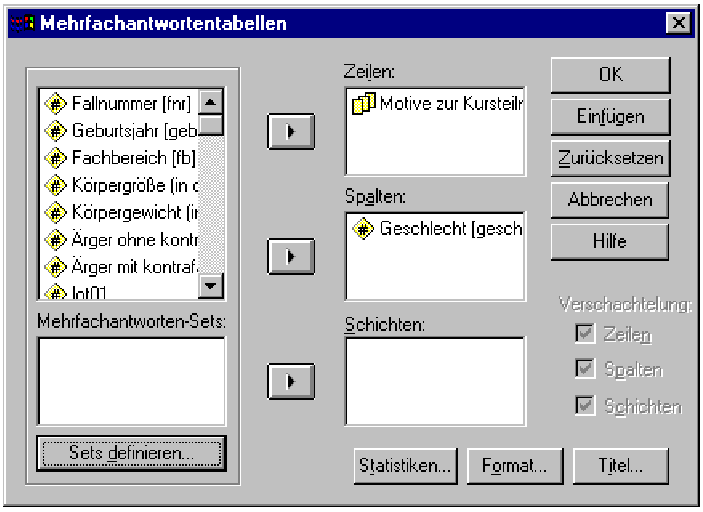
Wenn das SPSS-Zusatzmodul **Tables** zur Verfügung steht, können Kreuztabellen mit Mehrfachantworten-Sets auch im Pivot-Format erstellt werden:

14.3 Mehrfachantworten-Pivot-Tabellen

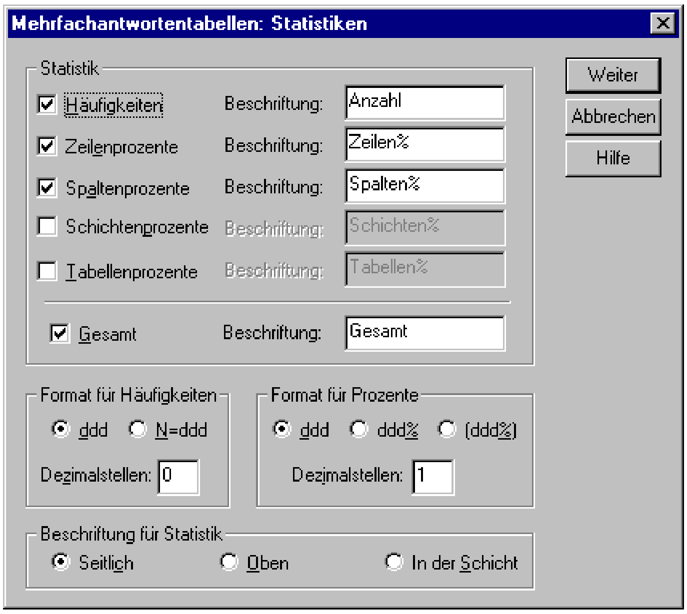
Leider produzieren die bisher beschriebenen Prozeduren für Mehrfachantworten-Sets keine Pivot-Tabellen, sondern nur Textausgaben. Wer über das SPSS-Zusatzmodul **Tables** verfügt, kann über den Menübefehl

Analysieren >Tabellen > Mehrfachantwortentabellen...

sowohl Häufigkeits- als auch Kreuztabellen für Mehrfachantworten-Sets in Pivot-Qualität erzeugen und dabei auch noch zusätzliche Gestaltungsmöglichkeiten nutzen:



Ergänzt man diese Dialogbox um folgende Angaben zu den gewünschten **Statistiken**



dann resultiert eine recht ansehnliche Pivot-Tabelle:

Geschlechtsunterschiede bei den Motiven zur Kursteilnahme

			Geschlecht		Gesamt
			Frau	Mann	
Motive zur Kursteilnahme	Eigene Studie	Anzahl	19	4	23
		Zeilen%	82,6	17,4	100,0
		Spalten%	79,2	66,7	76,7
	Bewerbung um Stelle	Anzahl	1		1
		Zeilen%	100,0		100,0
		Spalten%	4,2		3,3
	Bewerbung um HIWI-Job	Anzahl		1	1
		Zeilen%		100,0	100,0
		Spalten%		16,7	3,3
	Interesse an der EDV	Anzahl	3	2	5
		Zeilen%	60,0	40,0	100,0
		Spalten%	12,5	33,3	16,7
	Interesse an Statistik	Anzahl	9	1	10
		Zeilen%	90,0	10,0	100,0
		Spalten%	37,5	16,7	33,3
	Keine Aussage zutreffend	Anzahl	1		1
		Zeilen%	100,0		100,0
		Spalten%	4,2		3,3
Gesamt	Anzahl		24	6	30
	Zeilen%		80,0	20,0	100,0
	Spalten%		100,0	100,0	100,0

Die in Abschnitt 14.1 kritisierte MD-Konzeption der SPSS-Mehrfachantworten-Auswertung wird auch in den Pivot-Tabellen angewendet.

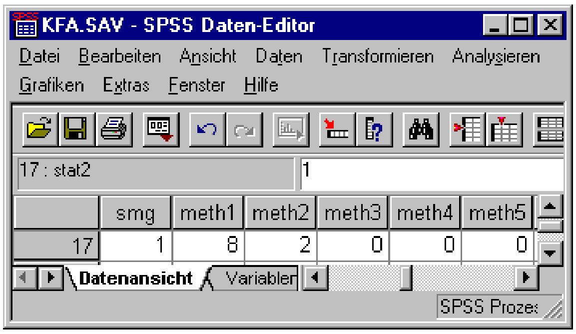
14.4 Ein sparsames Set kategorialer Variablen expandieren

In Abschnitt 1.4.2.3 wurde das sparsame Set aus kategorialen Variablen für Mehrfachwahl-Fragen mit sehr vielen Antwortmöglichkeiten als „Komprimierungsmethode“ empfohlen, um den Erfassungsaufwand gering zu halten. Zwar ist diese Datenstruktur kein Nachteil bei den Analyseprozeduren, die in den Abschnitten 14.1 und 14.2 beschrieben wurden, doch sind Auswertungen denkbar, die ein vollständiges Set aus dichotomen Variablen erfordern. In dieser Situation kann man das sparsame Set mit Hilfe der SPSS-Kommandosprache „expandieren“. Die folgenden Kommandos erzeugen zu unseren Variablen METH1 bis METH5 die acht dichotomen Variablen STAT1 bis STAT8, die für jeweils eine bestimmte statistische Methode festhalten, ob sie genannt worden ist (Wert 1) oder nicht (Wert 0):

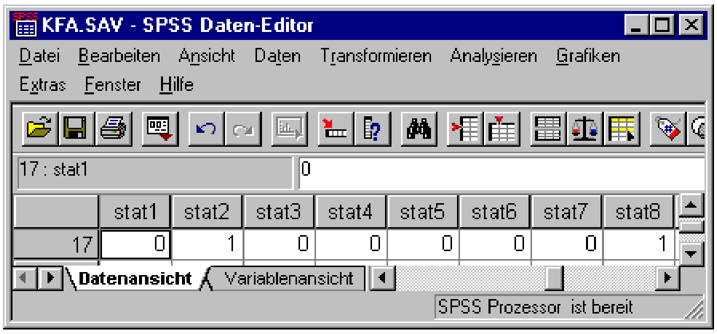
```
do repeat  stat = stat1 to stat8 /n = 1 to 8.
 do if (meth1 = n) or (meth2 = n) or (meth3 = n) or (meth4 = n) or (meth5 = n).
  compute stat = 1.
 else.
  compute stat = 0.
 end if.
end repeat.
execute.
```

Die Variable STAT2 steht z.B. für die Regressionsanalyse, weil gemäß Kodierplan bei einer der Variablen METH1 bis METH5 eine 2 zu notieren war, wenn ein Fall im Fragebogenteil 4b die Regressionsanalyse genannt hatte.

Beim Fall Nr. 17 wurden die genannten Methodenwünsche 8 (= logistische Regression) und 2 (= Regressionsanalyse) folgendermaßen mit dem sparsamen Set kategorialer Variablen METH1 bis METH5 erfasst:



Daraus ergeben sich folgende Werte für die Variablen STAT1 bis STAT8:



In obigem Programm werden zwei ausgesprochen nützliche Kontrollstrukturen der SPSS-Kommandosprache verwendet:

Schleife für strukturgleiche Transformationen

Die (DO REPEAT - END REPEAT) - Schleife wird achtmal ausgeführt, wobei im *i*-ten Umlauf die beiden Stellvertreter STAT und N gerade mit den *i*-ten Elementen der zugehörigen Listen identisch sind.

Fallunterscheidung

Beim Ausführen der (DO IF - ELSE - END IF) - Struktur passiert in Abhängigkeit vom Wahrheitswert des logischen Ausdruck mit den Variablen eines Falles folgendes:

Wert des logischen Ausdrucks	Aktion
wahr	Das erste COMPUTE-Kommando wird ausgeführt.
falsch	Das zweite COMPUTE-Kommando wird ausgeführt.
unbestimmt	Die neuen Variablen STAT1 bis STAT8 behalten den Initialisierungswert SYSMIS.

15 Datendateien im Textformat einlesen

Gelegentlich sind Daten auszuwerten, die in Text- bzw. ASCII-Dateien vorliegen. In Abschnitt 3.1.2 wurden zwei Datenformate beschrieben, die uns dabei begegnen können:

- positionierte Daten (feste Breite)
- separierte Daten (mit Trennzeichen).

Zum Importieren von Textdatendateien stellt SPSS einen leistungsfähigen Assistenten zur Verfügung, der mit

Datei > Textdaten einlesen

gestartet wird. Er kommt aber auch dann zum Einsatz, wenn Sie nach

Datei > Öffnen > Daten

eine Textdatendatei wählen.

Beim Einlesen einer (fehlerfreien) Datei mit separierten Werten sind kaum Probleme zu erwarten. Daher wird in diesem Abschnitt der etwas kniffligere Import von positionierten Daten beschrieben.

15.1 Import von positionierten Textdaten (feste Breite)

In der Datei **kfar-kv.txt**, die an der generell vereinbarten Stelle abgelegt ist (siehe Vorwort), finden Sie die gekoppelten KFA-Rohdaten aus drei anderen Kursen:

```
11 177115848
12  6 6 431214542432 110000
21 177115955
22  4 8 343335442442 110010
31 174416048
32  3 8 433224443342 100010
41 175116578
42  2 2 553125544531 100100
 .  .  .  .  .  .  .  .  .
 .  .  .  .  .  .  .  .  .
```

Es bietet sich an, diese Daten einzulesen, um die in Abschnitt 9 durch graphische Datenexploration gewonnene Moderator-Version der differentialpsychologischen Hypothese zu überprüfen.

Vom logischen Aufbau her besitzt auch die Datei **kfar-kv.txt** eine rechteckige Struktur, wobei die Daten eines Falles auf zwei Zeilen verteilt sind. Jede Variable hat eine feste Position im Datensatz eines Falles (z.B. Variable AERGO in Zeile 2, Spalten 5-6), so dass auch ihre Breite fixiert ist.
Die für uns relevanten Variablen haben folgende Positionen:

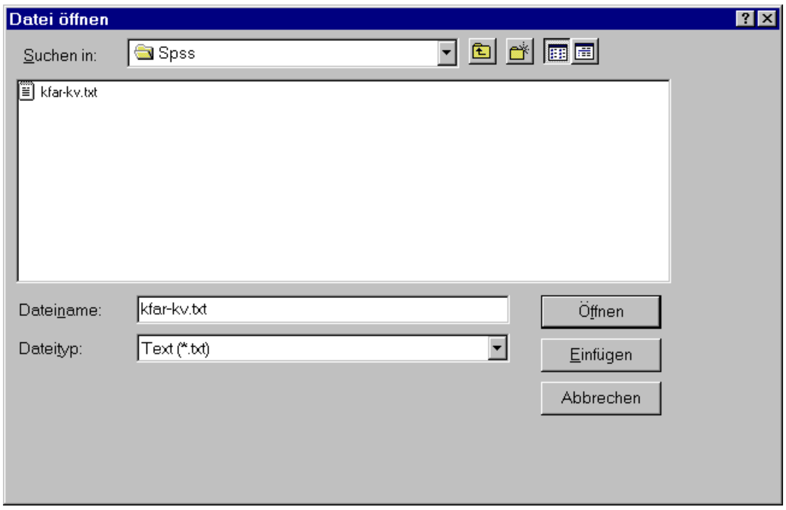
Variable	Datenzeile	Spalten
GESCHL	1	5
AERGO	2	5-6
AERGM	2	7-8
LOT01-LOT12	2	10-21

Alle übrigen Daten können wir ignorieren.

Gehen Sie folgendermaßen vor, um die relevanten Daten zu importieren:

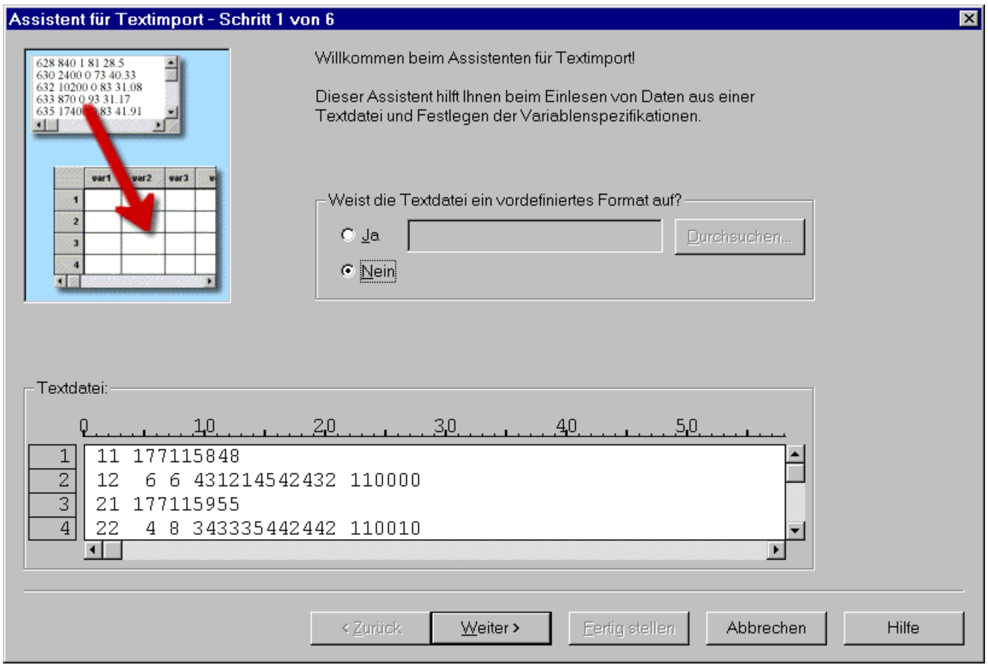
Textimport-Assistenten starten und Datei auswählen

Nach dem Start des Textimport-Assistenten ist zunächst die gewünschte Datei zu bestimmen, z.B.:



Schritt 1

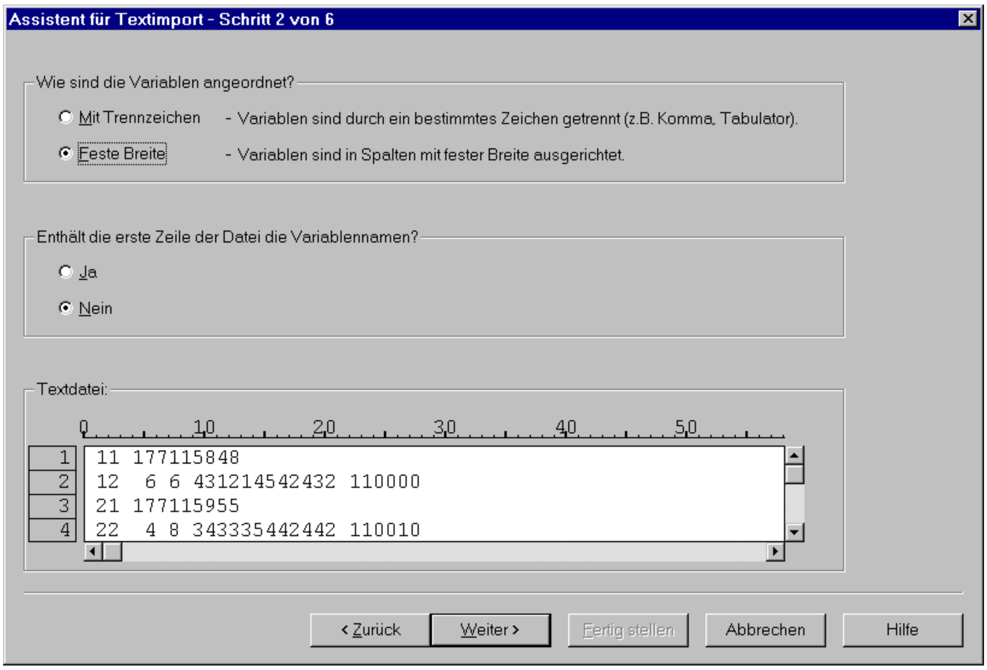
Im ersten Schritt zeigt der Assistent den Anfang unserer Datei und akzeptiert ggf. ein **vordefiniertes Format** aus früheren Assistenten-Einsätzen, das die Dateistruktur beschreibt.



Da wir auf eine solche Vorarbeit nicht zurückgreifen können, machen wir **weiter**.

Schritt 2

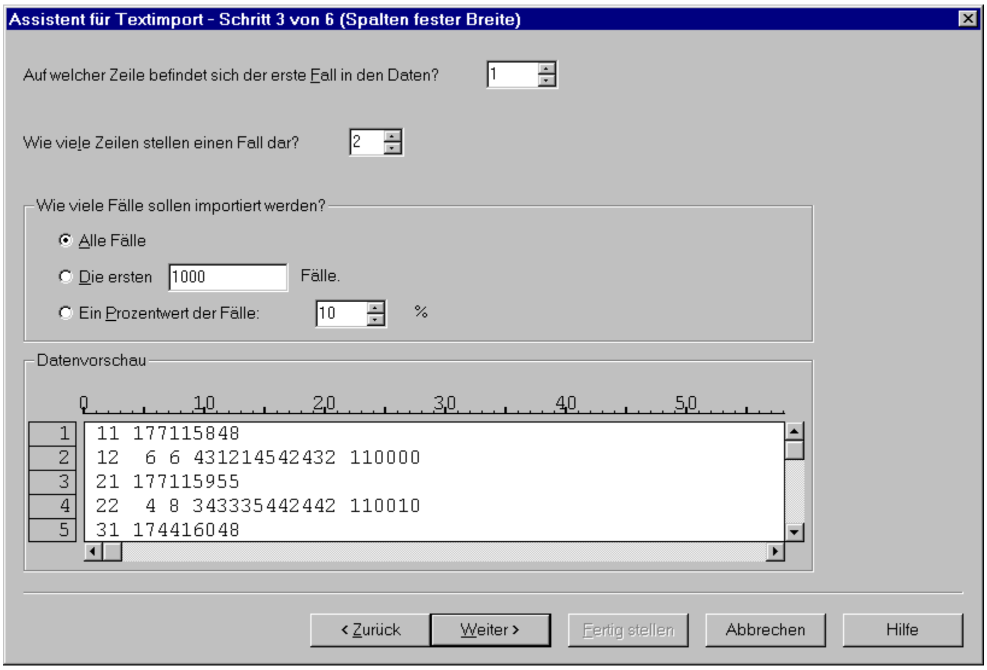
Im zweiten Schritt teilen wir mit, dass die Variablen in unserer Eingabedatei feste Positionen bzw. eine **feste Breite** besitzen:



Von der Möglichkeit, in der **ersten Zeile der Datei die Variablennamen** zu transportieren, wird in unserem Beispiel kein Gebrauch gemacht, was SPSS richtig erkannt hat.

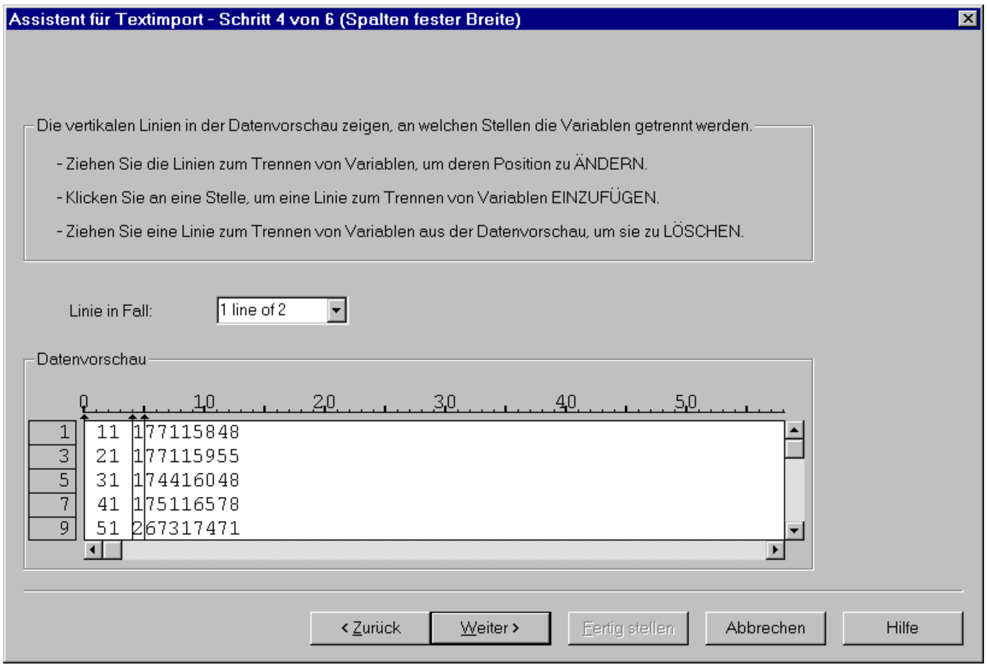
Schritt 3

Da unsere Datei keinen Vorspann enthält, **befindet sich der erste Fall** in Zeile 1. Der vom Text der Assistenten-Dialogbox diktierte letzte Satz ist nicht ganz korrekt, weil in unserem Beispiel jeweils zwei **Zeilen einen Fall darstellen**:

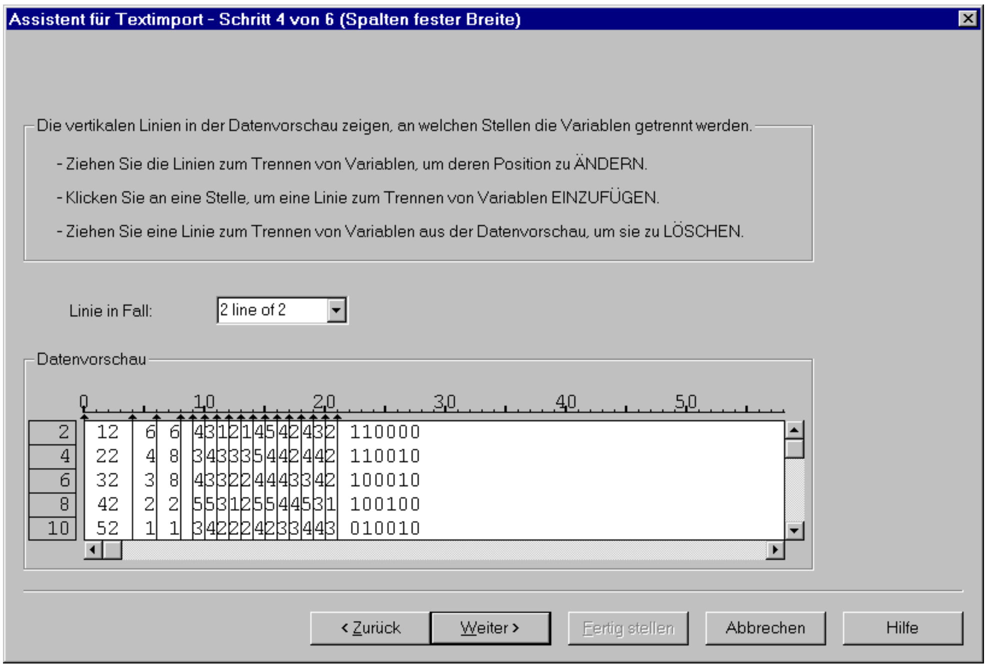


Schritt 4

Nun müssen wir die Positionen der einzulesenden Variablen festlegen, wobei der Assistent nur wenig Hilfestellung geben kann, wenn Variablen nicht separiert sind.
Von der ersten Datenzeile (**1 line of 2** in der Aufklapp-Liste **Linie in Fall**) benötigen wir nur die Variable GESCHL, die wir durch zwei senkrechte Linien abgrenzen:



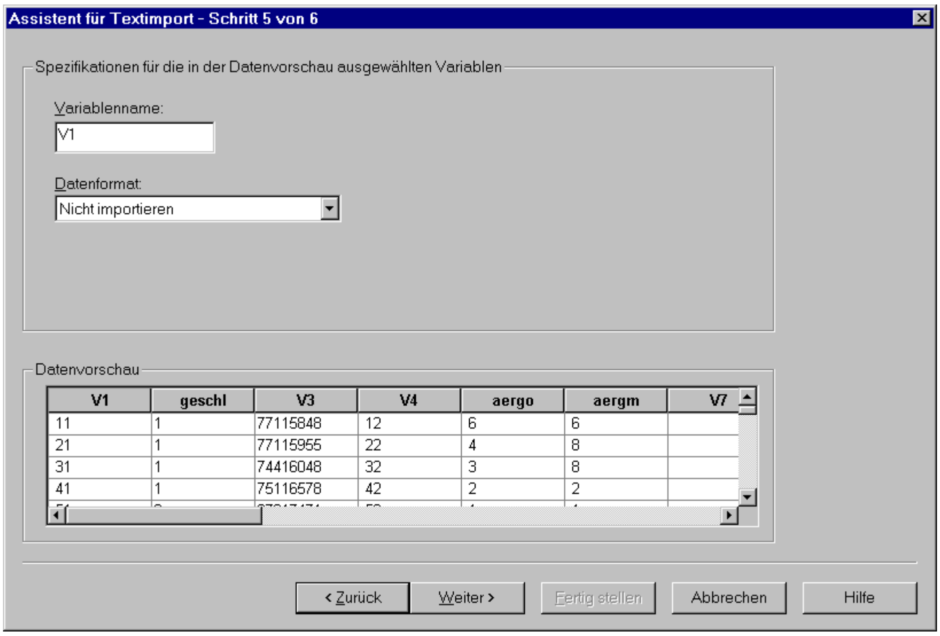
Wir die Trennlinien zu verwenden sind, ist in der Dialogbox beschrieben.
Auf der zweiten Datenzeile benötigen wir erheblich mehr Trennlinien:



Schritt 5

Im fünften Assistentenschritt können wir die von SPSS vorgeschlagenen Variablennamen ändern und ein **Datenformat** festlegen, worunter eine Typangabe mit zusätzlicher Formatspezifikation zu verstehen ist. Zum Umbenennen ist jeweils genau eine Spalte zu markieren. Vor einer *Format*spezifikation darf auch eine Variablen*liste* markiert werden.

Mit dem speziellen Datenformat **Nicht importieren** können überflüssige Variablen ausgeschlossen werden.



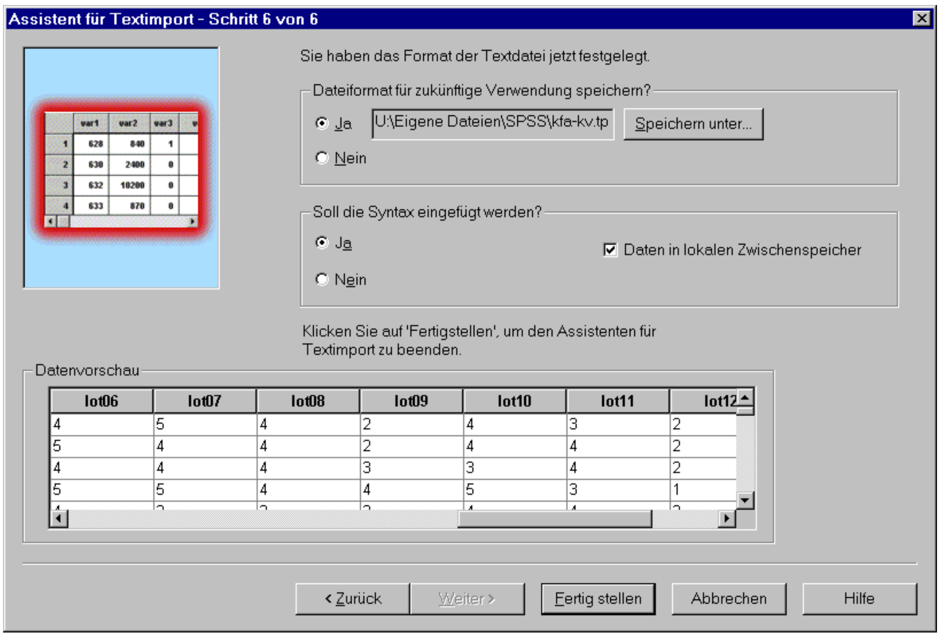
Zumindest bei den LOT-Variablen ist echte Fleißarbeit leisten, so dass wir nach Schritt 5 noch **weiter** machen, um unsere Arbeit zu konservieren.

Schritt 6

Der Assistent bietet zum Konservieren einer Dateispezifikation zwei Möglichkeiten:

- **Dateiformat für zukünftige Verwendung speichern?**
  Es entsteht eine Text-Assistenten-Format-Datei (Erweiterung .tpf), die bei einem späteren Assistenten-Einsatz im ersten Schritt angegeben werden kann (siehe oben).
- **Soll die Syntax eingefügt werden?**
  Das für den Datenimport verantwortliche GET DATA – Kommando wird in ein Syntaxfenster geschrieben. Es bietet sich an, zusätzliche Kommandos zu ergänzen, z.B. zum Deklarieren von MD-Indikatoren, die in den Textdaten vorhanden sind. Später kann mit Hilfe des entstandenen SPSS-Programms der Import mit allen erforderlichen Zusatzmaßnahmen automatisiert werden.

Es spricht nichts dagegen, beide Konservierungsoptionen zu verwenden:



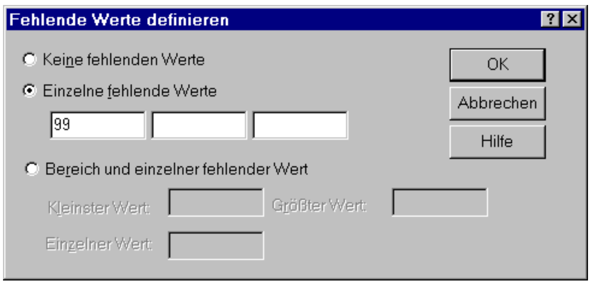
Das vom Textimport-Assistenten erzeugte Kommando GET DATA verblüfft mit einer wenig sinnvollen Spaltenzählung ab **0**, erfüllt aber ansonsten seinen Zweck:

```
GET DATA  /TYPE = TXT
 /FILE = 'U:\Eigene Dateien\SPSS\kfar-kv.txt'
 /FIXCASE = 2
 /ARRANGEMENT = FIXED
 /FIRSTCASE = 1
 /IMPORTCASE = ALL
 /VARIABLES =
 /1  V1 0-3 4X
 geschl 4-4 F1.0
 V3 5-12 8X
 /2 V4 0-3 4X
 aergo 4-5 F2.1
 aergm 6-7 F2.1
 V7 8-8 1X
 lot01 9-9 F1.0
 lot02 10-10 F1.0
 lot03 11-11 F1.0
 lot04 12-12 F1.0
 lot05 13-13 F1.0
 lot06 14-14 F1.0
 lot07 15-15 F1.0
 lot08 16-16 F1.0
 lot09 17-17 F1.0
 lot10 18-18 F1.0
 lot11 19-19 F1.0
 lot12 20-20 F1.0
 V20 21-27 7X
  .
CACHE.
EXECUTE.
```

Nach dem Einlesen einer ASCII-Datei dürfen Sie auf keinen Fall die Deklaration der dort eventuell verwendeten **MD-Indikatoren** vergessen. Studieren Sie also sorgfältig den hoffentlich vorhandenen Kodierplan, der in unserem Fall vorschreibt:

Variable	MD-Indikator
GESCHL	9
AERGO	99
AERGM	99
LOT01-LOT12	9

Die Deklaration kann in der Variablenansicht des Dateneditor erfolgen (siehe Abschnitt 3.2). Bei der Variablen AERGO ist z.B. für die Spalte **Fehlende Werte** einzutragen:



Das Kommando MISSING VALUES erlaubt allerdings eine rationellere MD-Deklaration:

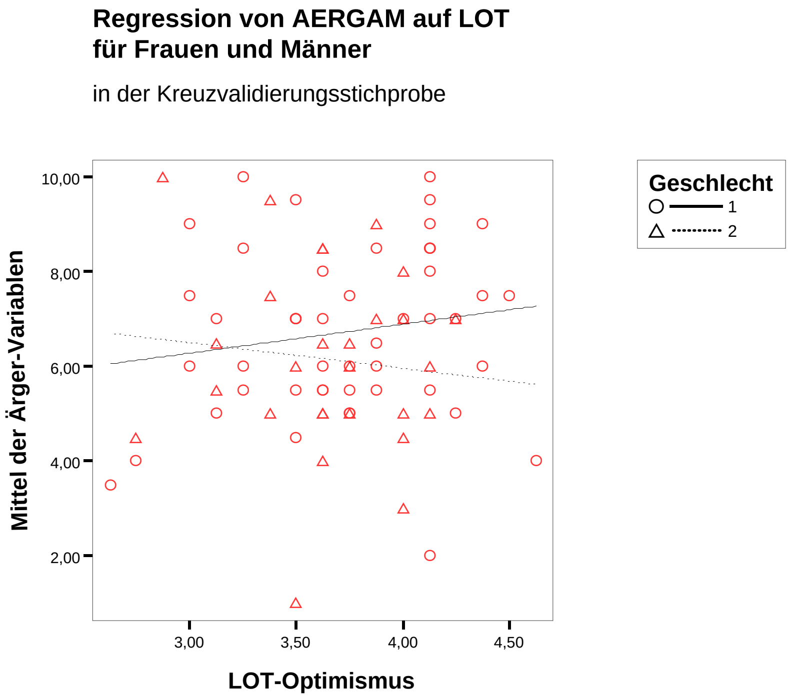
```
missing values  geschl (9) /aergo aergm (99) /lot01 to lot12 (9).
```

15.2 Überprüfung der revidierten differentialpsychologischen Hypothese

Um mit den neuen Daten die revidierte differentialpsychologische Hypothese prüfen zu können, sind zunächst einige Datentransformationen erforderlich, wobei wir uns die erforderlichen Kommandos aus dem Transformationsprogramm **kfat.sps** besorgen können:

```
* LOT-Fragen umkodieren.
RECODE
  lot03 lot04 lot05 lot12  (5=1)  (4=2)  (2=4)  (1=5)  .
EXECUTE .
* LOT berechnen.
COMPUTE lot = MEAN.6(lot01,lot03,lot04,lot05,lot08,lot09,lot11,lot12) .
VARIABLE LABELS lot 'LOT-Optimismus' .
EXECUTE .
* AERGAM berechnen.
COMPUTE aergam = (aergo + aergm)/2 .
VARIABLE LABELS aergam 'Mittel der Ärger-Variablen' .
EXECUTE .
```

Auch in der neuen Stichprobe scheint das Geschlecht die Regression von AERGAM auf LOT im erwarteten Sinn zu moderieren:



Allerdings wird der Interaktionseffekt nicht signifikant (p = 0,307):

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten B	Nicht standardisierte Koeffizienten Standardfehler	Standardisierte Koeffizienten Beta	T	Signifikanz
1	(Konstante)	,773	5,562		,139	,890
	Geschlecht	3,670	4,130	,949	,889	,377
	LOT-Optimismus	1,761	1,493	,413	1,180	,242
	GESCHL * LOT	-1,150	1,118	-1,120	-1,029	,307

a. Abhängige Variable: Mittel der Ärger-Variablen

Wer will, kann die differentialpsychologische Hypothese nach der neuerlichen Pleite nun noch weiter exhaurieren.
Am Ende des Kurses will ich Ihnen lieber gestehen, dass diese Hypothese vermutlich ein Nonsens ist und im Kurs lediglich dazu diente, den Umgang mit „gescheiterten“ Hypothesen zu demonstrieren.

16 Anhang

16.1 Kurzbeschreibung der SPSS-Module und Literaturhinweise

SPSS für Windows besteht aus einem Basissystem und mehreren optionalen Erweiterungsmodulen. An dieser Stelle möchte ich die Leser(innen) nicht mit einer Beschreibung der einzelnen Programmbestandteile langweilen, doch einige Informationen erscheinen sinnvoll. Vielleicht müssen Sie z.B. einmal beurteilen, ob mit den bei einer konkreten SPSS-Installation vorhandenen Modulen eine Zeitreihenanalyse möglich ist.

Die vielfältigen SPSS-Analysemöglichkeiten sind in zahlreichen Handbüchern beschrieben, die wertvolle Informationen zur Bedienung des Programms und zu den angebotenen statistischen Methoden enthalten. Im folgenden werden zu den SPSS-Modulen die Original-Handbücher sowie einige weitere Publikationen angegeben.
In diesem Zusammenhang soll nochmals an eine oft vernachlässigte, durchaus brauchbare, vollständige, stets verfügbare und kostenlose Darstellung von SPSS für Windows erinnert werden: an das Online-Hilfesystem des Programms.

Neben den Erweiterungsmodulen bietet SPSS noch mehrere **Zusatzprodukte** an, welche die Datenerfassung unterstützen oder spezielle Analysemöglichkeiten bieten (z.B. Data Entry, Amos, AnswerTree, Neural Connection). Es handelt sich um selbständig einsetzbare Programme, die aber gut mit SPSS kooperieren.
In diesem Manuskript wurden die Zusatzprodukte nicht behandelt, bestenfalls erwähnt. Im Rahmen von Rechenzentrums-Kursen sind etliche Bedienungsanleitungen zu den Zusatzprodukten oder Ausarbeitungen zu den dort unterstützten statistischen Methoden entstanden, die Sie als elektronische Dokumente auf dem WWW-Server der Universität Trier von der Startseite (www.uni-trier.de) ausgehend folgendermaßen finden:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen

16.1.1 Base System

Das Basispaket bietet u.a. folgende Funktionen:

- Datenmanagement und -transformation, Dateneditor
- Grundlegende Auswertungsverfahren, z.B.:
    - Darstellung von Verteilungen
    - Kreuztabellen-Analyse
    - Mittelwertsvergleiche
    - Korrelationen (auch partiell)
    - Lineare Regression
    - Reliabilitätsanalyse
    - mehrdimensionale Skalierung
    - Mehrfachwahl-Analyse
    - Einfache varianzanalytische Auswertungen (faktorielle Designs)
    - Nonparametrische Statistik
    - Diskriminanzanalyse
    - Faktorenanalyse
    - Clusteranalyse
    - Reporterstellung
- Umfangreiche graphische Darstellungsmöglichkeiten

In englischer Sprache bietet SPSS folgende Handbücher an:

- *SPSS Base 10.0. User's Guide Package*
  Preis: DM 107,-
  Das Paket besteht aus den beiden folgenden Bänden:
  - SPSS Inc. (1999). *SPSS Base 10.0 User's Guide*. Prentice Hall.
    537 Seiten, ISBN: 0-13-017902-7
    Das Buch beschreibt die Benutzeroberfläche, grundlegende Optionen zur Behandlung von Daten und SPSS-Ergebnissen, Bedienungshilfen sowie Konfigurationsmöglichkeiten. Neben den EDV-technischen Aspekten der SPSS-Anwendung werden auch die statistischen Analysemöglichkeiten des Basissystems beschrieben, allerdings mit relativ knappen Angaben zu den verfügbaren statistischen Methoden.
  - SPSS Inc. (1999). *SPSS Interactive Graphics 10.0*. Prentice Hall.
    261 Seiten, ISBN: 0-13-017903-5
    In diesem Buch werden die über den Menübefehl
    Grafiken > Interaktiv
    verfügbaren Darstellungsmöglichkeiten beschrieben.
    Die interaktive Grafik bietet mehr Flexibilität und Gestaltungsmöglichkeiten als die ältere, parallel weiterhin vorhandene, statische Grafik. Viele Diagrammtypen werden jedoch nur vom Modul für statischen Grafiken unterstützt.
- SPSS Inc. (1999). *SPSS Base 10.0 Applications Guide*. Prentice Hall.
  426 Seiten, ISBN: 0-13-017901-9, Preis: DM 107,-
  Hier werden die Prozeduren des Basissystems anwendungsorientiert anhand vollständig diskutierter Beispiele beschrieben, wobei auch die verwendeten statistischen Methoden erläutert werden.
- SPSS Inc. (1999). *SPSS 10.0 Syntax Reference guide*. Prentice Hall.
  1433 Seiten, Preis: DM 119,-
  Das Buch ist als PDF-Dokument in das SPSS-Hilfesystem integriert (**Hilfe > Syntax Guide Base**), so dass eine Anschaffung in gedruckter Form kaum erforderlich ist. Es beschreibt im wesentlichen die zu den Modulen Base, Regression Models und Advanced Models gehörigen SPSS-Kommandos, enthält aber auch Literaturhinweise zu den statistischen Verfahren.
- Norusis, M. J. / SPSS Inc. (2000). *SPSS 10.0 Guide to Data Analysis.* Prentice Hall.
  577 Seiten, Preis: DM 119,-, ISBN: 0-13-029204-4
  Dieses sehr empfehlenswerte Werk kann annähernd als ein Lehrbuch der empirischen Forschung und der Statistik mit starker Berücksichtigung des Werkzeugs SPSS charakterisiert werden.

In deutscher Sprache bietet SPSS das folgende Handbuch an:

SPSS Inc. (1999). *SPSS Base 10.0 Benutzerhandbuch*. München: SPSS GmbH Software.
591 Seiten, ISBN: 1-56827-853-5, Preis: DM 107,- (inkl. Interactive Graphics)

Hier handelt es sich um eine Übersetzung des *SPSS Base 10.0. User's Guide*.

Auch in anderen Verlagen sind mittlerweile zahlreiche deutschsprachige Titel zur statistischen Datenanalyse mit SPSS erschienen. Besonders interessant erscheinen mir Bücher, die statistische Methoden grundlegend oder anwendungsorientiert darstellen und in den Beispielen zur Erläuterung der erklärten Begriffe bzw. Verfahren auch kommentierte SPSS-Ausgaben bieten, z.B.:

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2000). *Multivariate Analysemethoden* (9. Aufl.). Berlin: Springer.
660 Seiten, ISBN 3-540-67146-3, Preis: DM 65,-

Ein URT-Kursmanuskript zur *linearen Regression* mit SPSS 8 finden Sie im Internet von der Startseite der Universität Trier (www.uni-trier.de) ausgehend unter:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen > Statistische Spezialthemen > Lineare Regressionsanalyse mit SPSS

16.1.2 Regression Models

Es sind u.a. die folgenden statistischen Verfahren enthalten:

- Logistische Regression (auch multinomial)
- Probit-Analyse
- Nichtlineare Regression
- WLS- und TSLS-Regression

Von SPSS wird in englischer Sprache das folgende Handbuch angeboten:

Norusis, M. J. & SPSS Inc. (1999). *SPSS Regression Models 10.0*. Prentice Hall.
135 Seiten, ISBN: 0-13-017904-3, Preis: DM 107,-
Es erklärt die Prozeduren im selben Stil wie der *SPSS Base 10.0 Applications Guide*.

Zur logistischen sowie zur nichtlinearen Regressionsanalyse mit SPSS finden Sie URT-Kursmanuskripte im Internet von der Startseite der Universität Trier (www.uni-trier.de) ausgehend unter:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen > Statistische Spezialthemen

16.1.3 Advanced Models

Das Modul enthält u.a. die folgenden Verfahren:

- Multivariate Varianzanalyse (inklusive Meßwiederholungsmodelle)
- Varianzkomponentenzerlegung
- Log- und Logitlineare Analyse
- Regressionsanalyse für ordinale Kriterien
- Sterbetafelanalyse (inkl. Kaplan-Meier)
- Cox-Regression

Von SPSS wird in englischer Sprache folgendes Handbuch angeboten:

SPSS Inc. (1999). *SPSS Advanced Models 10.0*. Prentice Hall.
333 Seiten, ISBN: 0-13-017890-X, Preis: DM 107,-
Es erklärt die Prozeduren im selben Stil wie der *SPSS Base 10.0 Applications Guide*.

Es existiert auch eine ins Deutsche übersetzte Version des Handbuchs. Leider wurden dabei die ausgearbeiteten Beispiele weggelassen, auf denen die Attraktivität des Originals wesentlich beruht:

SPSS Inc. (1999). *SPSS Advanced Models 10.0*. München: SPSS GmbH Software.
125 Seiten, ISBN: 1-56827-858-6, Preis: DM 83,-

16.1.4 Tables

Dieses Modul erlaubt die Erstellung frei gestalteter Tabellen, wobei Sie nicht nur das Design, sondern in weiten Grenzen auch die Inhalte bestimmen können.

Das englischsprachige SPSS-Handbuch zu **Tables**:

> SPSS Inc. (1998). *SPSS Tables 8.90*. Prentice Hall.
> 228 Seiten, ISBN: 1-56827-210-3, Preis: DM 84,-

16.1.5 Trends

Trends beherrscht zahlreiche Auswertungsmethoden aus dem Bereich der Zeitreihenanalyse, z.B. ARIMA-Modellierung, Spektralanalyse, verschiedene Prognosemethoden, regressionsanalytische Methoden für Zeitreihen.

Das SPSS-Handbuch zu **Trends**:

> SPSS Inc. (1999). *SPSS Trends10.0*. Prentice Hall.
> 307 Seiten, ISBN: 0-13-017905-1, Preis: DM 83,-

Ein URT-Kursmanuskript zu **Trends** finden Sie im Internet von der Startseite der Universität Trier (www.uni-trier.de) ausgehend unter:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen > Statistische Spezialthemen > Zeitreihenanalyse mit SPSS-Trends

16.1.6 Categories

Categories unterstützt verschiedene Verfahren zur Analyse kategorialer Daten:

- Regression mit optimaler Skalierung
- Nichtlineare Hauptkomponentenanalyse
- Nichtlineare kanonische Korrelation
- Korrespondenzanalyse
- Homogenitätsanalyse.
- Multidimensionale Skalierung

SPSS bietet in englischer Sprache das folgende Handbuch an:

> Meulman, J. J., Heiser, W. J. & SPSS Inc. (1999). *SPSS Categories 10.0*. Chicago, IL: SPSS Inc.
> 322 Seiten, Preis: DM 95,-

Es existiert auch eine ins Deutsche übersetzte Version des Handbuchs. Leider wurden dabei die ausgearbeiteten Beispiele weggelassen, auf denen die Attraktivität des Originals wesentlich beruht:

> Meulman, J. J. & Heiser, W. J. & SPSS Inc. (1999). *SPSS Categories 10.0*. München: SPSS GmbH Software.
> 99 Seiten, ISBN: 1-56827-863-2, Preis: DM 48,-

Zur Korrespondenz- und zur Homogenitätsanalyse mit SPSS finden Sie URT-Kursmanuskripte im Internet von der Startseite der Universität Trier (www.uni-trier.de) ausgehend unter:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen > Statistische Spezialthemen

16.1.7 Conjoint

Dieses Modul unterstützt die Marktforschung per Conjoint-Analyse.

SPSS bietet in englischer Sprache das folgende Handbuch an:

SPSS Inc. (1997). *SPSS Conjoint 8.0*. Prentice Hall.
76 Seiten, ISBN: 1-56827-212-X, Preis: DM 46,-

Ein URT-Kursmanuskript zur Conjoint-Analyse mit SPSS finden Sie im Internet von der Startseite der Universität Trier (www.uni-trier.de) ausgehend auf folgendem Weg:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen >
Statistische Spezialthemen > Conjoint-Analyse mit SPSS 6.1

16.1.8 Exact Tests

Zahlreiche klassische Signifikanztests, die in modernen Statistik-Programmpaketen (auch in SPSS) durchweg verwendet werden, gelten nur *approximativ* und dürfen daher bei vielen kleinen Stichproben nicht benutzt werden. Diese Einschränkungen werden im Modul **Exact Tests** für die Kreuztabellenanalyse und etliche Verfahren der nonparametrischen Statistik überwunden.

Das SPSS-Handbuch zu **Exact Tests**:

Metha, C.R., Patel, N.R. & SPSS Inc. (1996). *Exact Tests 7.0 for Windows*. Prentice Hall.
220 Seiten, ISBN: 1-56827-108-5

Ein URT-Kursmanuskript zu **Exact Tests** finden Sie im Internet von der Startseite der Universität Trier (www.uni-trier.de) ausgehend auf folgendem Weg:

Weitere Serviceangebote > EDV-Dokumentationen > Elektronische Publikationen >
Statistische Spezialthemen > Exakte Tests mit SPSS

16.1.9 Informations- und Bezugsquellen zu den SPSS-Handbüchern

16.1.9.1 In Trier verfügbare SPSS-Handbücher

Welche SPSS-Handbücher in Trierer Bibliotheken verfügbar sind, können Sie z.B. über eine Suche im Online-Katalog der Universitätsbibliothek in Erfahrung bringen.

Die SPSS-Originalhandbücher stehen auch in der Benutzerberatung des Rechenzentrums zur Einsicht und Kurzausleihe bereit.

16.1.9.2 Zum Bezug der SPSS-Originalhandbücher

Die SPSS-Originalhandbücher werden vom Verlag **Prentice Hall** vertrieben und sind im Buchhandel frei erhältlich. Als deutscher Vertriebspartner wurde von Prentice Hall folgende Direktversand-Buchhandlung autorisiert:

BSB Distribution
Andernacher Straße 46
D-56070 Koblenz
Tel. 0261/98256-0
Fax 98256-16
Internet: www.bsb.de

Beim Erwerb von SPSS-*Original*-Handbüchern über diese Buchhandlung sollte die Universität Trier als Inhaberin einer SPSS-Mehrfachlizenz einen Rabatt von 25% erhalten.

16.2 Weitere Hinweise zur SPSS-Kommandosprache

Im Hauptteil des Manuskriptes wurden nur sehr oberflächliche Hinweise zur SPSS-Kommandosprache gegeben. Diese sollten genügen für Anwender(innen), die nicht frei programmieren, sondern nur gelegentlich ein von SPSS automatisch erzeugtes Kommando modifizieren wollen.
Der aktuelle Abschnitt ist für ambitionierte Anwender gedacht, die bereit sind, SPSS-Programme zu schreiben, ...

- um auch die nur per Syntax verfügbaren SPSS-Leistungen nutzen zu können,
- um rationeller mit SPSS arbeiten zu können.

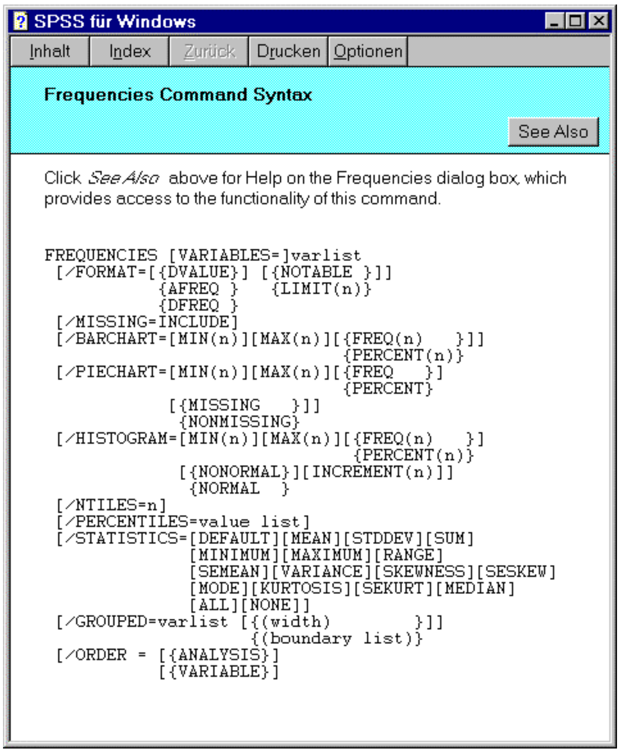
16.2.1 Hilfsmittel für das Arbeiten mit der SPSS-Kommandosprache

Die wichtigsten Hilfsmittel sind die **Syntaxhandbücher**, die SPSS freundlicherweise dem Programm in elektronischer Form (als PDF-Dokumente) beigepackt und über das Hilfesystem zugänglich gemacht hat:

Hilfe > Syntax Guide

In den Syntaxhandbüchern zu den installierten Modulen finden Sie ausführliche Beschreibungen der zum jeweiligen Modul gehörigen SPSS-Kommandos mit zahlreichen kommentierten Beispielen. Für die Module Base, Regression Models und Advanced Models hat SPSS ein gemeinsames Syntaxhandbuch erstellt, das unter dem Titel *Base* zu finden ist.

Die Syntaxfenster bieten ein einfaches Verfahren, das **Syntaxdiagramm** zu einem konkreten Kommando einzusehen: Setzen Sie die Schreibmarke auf ein Beispiel des Kommandos, und klicken Sie dann auf das Symbol . Zum FREQUENCIES-Kommando, das der **Häufigkeiten**-Dialogbox zugrunde liegt, erscheint z.B. das folgende Hilfefenster:



SPSS für Windows

Inhalt | Index | Zurück | Drucken | Optionen

Frequencies Command Syntax

See Also

Click *See Also* above for Help on the Frequencies dialog box, which provides access to the functionality of this command.

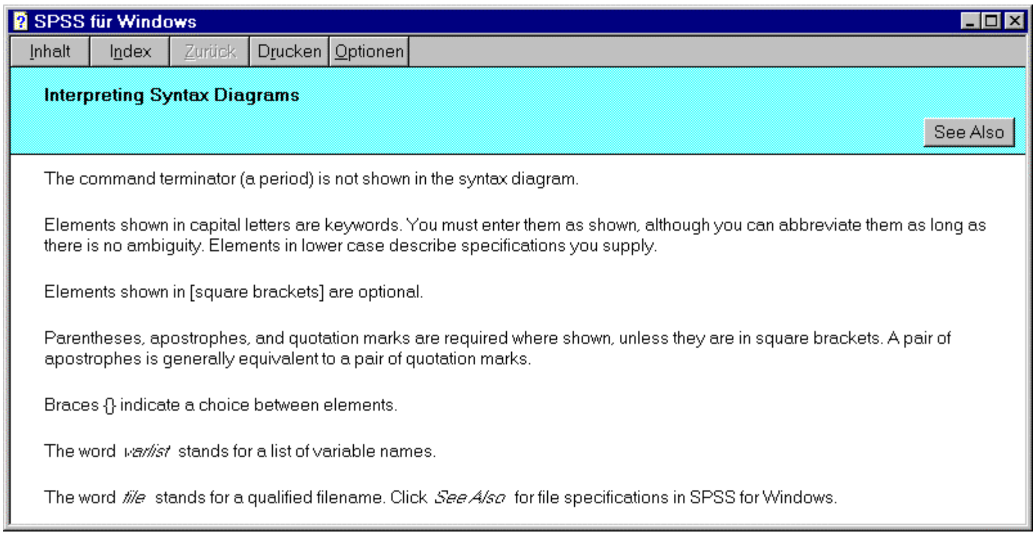
```
FREQUENCIES [VARIABLES=]varlist
 [/FORMAT=[{DVALUE}] [{NOTABLE }]]
           {AFREQ }   {LIMIT(n)}
           {DFREQ }
 [/MISSING=INCLUDE]
 [/BARCHART=[MIN(n)][MAX(n)][{FREQ(n)   }]]
                             {PERCENT(n)}
 [/PIECHART=[MIN(n)][MAX(n)][{FREQ   }]
                             {PERCENT}
            [{MISSING   }]]
             {NONMISSING}
 [/HISTOGRAM=[MIN(n)][MAX(n)][{FREQ(n)   }]
                              {PERCENT(n)}
             [{NONORMAL}][INCREMENT(n)]]
              {NORMAL  }
 [/NTILES=n]
 [/PERCENTILES=value list]
 [/STATISTICS=[DEFAULT][MEAN][STDDEV][SUM]
              [MINIMUM][MAXIMUM][RANGE]
              [SEMEAN][VARIANCE][SKEWNESS][SESKEW]
              [MODE][KURTOSIS][SEKURT][MEDIAN]
              [ALL][NONE]]
 [/GROUPED=varlist [{(width)        }]]
                    {(boundary list)}
 [/ORDER = [{ANALYSIS}]
           [{VARIABLE}]
```

16.2.2 Interpretation von Syntaxdiagrammen

Mit dem Syntaxdiagramm wird die allgemeine Form des Kommandos definiert und somit festgelegt, wie konkrete Beispiele gebildet werden müssen. Solche Syntaxdiagramme werden auch im weiteren Verlauf dieses Abschnitts benutzt, um Bestandteile der SPSS-Sprache zu erläutern. In den Syntaxdiagrammen treten einige Metazeichen auf (z.B. "[", "{"), die nicht zur SPSS-Sprache selbst gehören, sondern diese Sprache beschreiben. Die Bedeutung dieser Metazeichen müssen Sie kennen, um Syntaxdiagramme richtig interpretieren zu können. Im Hilfesystem finden Sie eine Erklärung, indem Sie nach

Hilfe > Themen > Index

den Suchbegriff „Syntax“ in das aktive Textfeld eintippen und dann einen Doppelklick auf den Eintrag **syntax rules** setzen:



SPSS für Windows

Inhalt | Index | Zurück | Drucken | Optionen

Interpreting Syntax Diagrams

See Also

The command terminator (a period) is not shown in the syntax diagram.

Elements shown in capital letters are keywords. You must enter them as shown, although you can abbreviate them as long as there is no ambiguity. Elements in lower case describe specifications you supply.

Elements shown in [square brackets] are optional.

Parentheses, apostrophes, and quotation marks are required where shown, unless they are in square brackets. A pair of apostrophes is generally equivalent to a pair of quotation marks.

Braces {} indicate a choice between elements.

The word *varlist* stands for a list of variable names.

The word *file* stands for a qualified filename. Click *See Also* for file specifications in SPSS for Windows.

16.2.3 Aufbau eines SPSS-Programms

Welche Kommandos SPSS für das Erstellen von Programmen bereithalten muss, ergibt sich aus unseren Zielvorstellungen: Wir möchten SPSS anweisen, unsere empirischen Daten zu lesen, gegebenenfalls aus den gelesenen Variablen interessantere neue Variablen zu berechnen und schließlich statistische Verfahren mit den eingelesenen oder neu erstellten Variablen zu rechnen. Darüber hinaus haben wir gelegentlich Sonderwünsche hinsichtlich der Arbeitsweise von SPSS.

Orientiert an den gerade skizzierten Teilaufgaben unserer Auswertungsaufträge kann man die verfügbaren SPSS-Kommandos in folgende Gruppen einteilen:

- **Dateidefinitions-Kommandos**
  Sie dienen zum Einlesen von Daten in die Arbeitsdatei. Als Beispiel haben wir bereits das GET-Kommando kennen gelernt.
  Wenn ein Programm kein Dateidefinitions-Kommando enthält, wenn es also nicht selbst für das Einlesen seiner Daten sorgt, kann es natürlich nur ausgeführt werden, wenn zuvor eine Arbeitsdatei erzeugt worden ist.
- **Transformations-Kommandos**
  Diese Kommandos dienen zur Veränderung oder Neuberechnung von Variablen bzw. zur Auswahl von Fällen für die weitere Verarbeitung.
- **Prozedur-Kommandos**
  Damit werden statistische Analysen, graphische Präsentationen oder Dateibearbeitungen (z.B. Sortieren der Fälle) angefordert. Ein Beispiel ist das FREQUENCIES-Kommando.
- **Dienst-Kommandos**
  Damit kann man u.a. die Arbeitsweise von SPSS beeinflussen (z.B. Startwert des Pseudozufallszahlengenerators setzen) und verschiedene Informationen anfordern.

In folgendem SPSS-Programm treten Kommandos aus allen Gruppen auf:

```
comment Größe und Gewicht.                    | Dienst-Kommando

get file = 'kfar.sav'.                        | Dateidef.-Kommando

frequencies var = groesse gewicht             | Prozedur-
 /statistics = all /histogram = normal.       | Kommando

compute  ideal = groesse - 100.               | Transformations-
                                              | Kommando

t-test  pairs = gewicht ideal.                | Prozedur-
                                              | Kommando
```

SPSS-Programme können flexibel gestaltet werden:

- Für die Reihenfolge der SPSS-Kommandos gilt lediglich die selbstverständliche Regel, dass auf eine Variable erst dann Bezug genommen werden darf, nachdem sie im Rahmen einer Dateidefinition oder durch ein Transformations-Kommando eingeführt worden ist.

- In einem Programm dürfen beliebig viele Prozedur-Kommandos auftreten.
  Manche Anwender leben in dem Irrglauben, pro SPSS-Programm sei nur eine einzige Statistik-Prozedur erlaubt, und verstreuen daher zusammenhängende Auswertungen über unübersichtlich viele Mini-Programme.
  Andere haben den falschen Ehrgeiz, ihr gesamtes Projekt in einem einzigen Programm abzuwickeln, und erstellen dabei ein unpraktisches Monster-Programm mit mehreren hundert Zeilen.
  Wie so oft im Leben ist auch hier der gesunde Mittelweg zu empfehlen: Eng zusammengehörige Auswertungsschritte sollten in einem Programm gesammelt werden (z.B. alle Prozeduren zur Datenprüfung), für abgrenzbare Aufgabenpakete bzw. Analysezusammenhänge sollte jeweils ein eigenes Programm erstellt werden.
- Auch nach einer Prozedur dürfen Daten-Transformationen vorgenommen werden.
- Man kann nach einer Prozedur sogar weitermachen mit der Definition einer neuen Arbeitsdatei, welche dann die alte ersetzt.

16.2.4 Aufbau eines einzelnen SPSS-Kommandos

Die wichtigsten Regeln für SPSS-Befehle:

- Ein Kommando besteht aus seinem Namen und den zugehörigen Spezifikationen:

kommandoname spezifikationen

  - Der **Kommandoname** kann aus einem Wort bestehen oder aus mehreren Wörtern.
    Beispiele:
    - FREQUENCIES
    - GET DATA
  - Die **Spezifikationen** dürfen enthalten:
    - Schlüsselwörter (z.B. VARIABLES)
    - Variablennamen
    - Zahlen
    - Zeichenfolgen (z.B. Variablenetiketten)
    - Operatoren (z.B. "+")
    - spezielle Begrenzungszeichen: / ( ) = ' "

    Zwischen diesen Elementen ist mindestens ein Leerzeichen erforderlich. Ausnahme:
    Die speziellen Begrenzungszeichen, die arithmetischen Operatoren und manche Vergleichsoperatoren (z.B. ">") sind selbstbegrenzend, d.h. vor und nach ihnen sind keine Leerzeichen nötig (aber erlaubt).
    Statt eines Leerzeichens darf man meist verwenden:
    - beliebig viele Leerzeichen,
    - ein Komma,
    - einen Zeilenwechsel.

    Dies ermöglicht eine übersichtliche Programmgestaltung.

- *Innerhalb* eines Kommandos sind keine Leerzeilen erlaubt.

- Jedes Kommando muss in einer neuen Zeile beginnen und mit einem Punkt enden.
  Die Kommandos müssen dabei keinesfalls in der ersten *Spalte* beginnen, sondern dürfen eingerückt werden. Von dieser Möglichkeit sollte man z.B. bei Schleifen-Konstruktionen Gebrauch machen.

Beispiel:
```
do repeat  mc=mc001 to mc100.
  compute  mc=normal(1).
end repeat.
```
Hier werden 100 unabhängige, normalverteilte Zufallsvariablen erzeugt. Durch das Einrücken wird deutlich gemacht, dass die COMPUTE-Anweisung innerhalb der DO REPEAT - Schleife steht.

- In SPSS für Windows brauchen Sie keine maximale Länge für Programmzeilen zu beachten. Manche andere SPSS-Versionen, unter denen Ihr Programm möglicherweise auch laufen soll, haben jedoch eine Beschränkung auf 80 Spalten.

- Ein Kommando kann sich über beliebig viele Fortsetzungszeilen erstrecken.

- Die Verwendung von Groß- oder Kleinbuchstaben ist beliebig.

- Schlüsselwörter dürfen meist bis auf die ersten drei Zeichen abgekürzt werden.
  Beispiel: "`fre`" für "`frequencies`"

- Bei den meisten Kommandos sind die Spezifikationen in Subkommandos unterteilt. Diese beginnen mit einem Subkommando-Namen, meist gefolgt von einem Gleichheitszeichen, und sind durch Schrägstriche voneinander getrennt.
  Beispiel:
  ```
  frequencies var=lot01 /format=notable
    /statistics=all.
  ```

Merken Sie sich bitte aus dieser Liste für den Anfang vor allem:

JEDES KOMMANDO MUSS IN EINER NEUEN ZEILE BEGINNEN UND MIT EINEM PUNKT ENDEN.

16.2.5 Regeln für Variablenlisten

16.2.5.1 Abkürzende Spezifikation einer Serie von Variablen

In Transformations- oder Prozedur-Kommandos soll häufig eine Folge **bereits existierender** und **in der Arbeitsdatei hintereinander liegender** Variablen angesprochen werden. Dies ermöglicht das **aufrufende TO**, dessen Syntax im folgenden erläutert wird:

vara TO *varb*

vara, *varb*	Namen bereits vorhandener Variablen, wobei *vara* in der Arbeitsdatei vor *varb* stehen muß.

Beispiele:

- `frequencies var=alter to beruf.`
  Für alle Variablen, die in der Arbeitsdatei zwischen ALTER und BERUF positioniert sind, werden Häufigkeitstabellen erstellt.
- `frequencies var=frage1 to frage3.`
  Wenn in der Arbeitsdatei zwischen FRAGE1 und FRAGE3 1500 beliebig benannte Variablen stehen, dann bewirkt das Kommando 1502 Häufigkeitstabellen.

16.2.5.2 Der Platzhalter „varlist"

In folgendem Syntaxdiagramm wird der in SPSS-Kommandos häufig auftretende Platzhalter *varlist* definiert:

{*varname* | *varname_1* TO *varname_2*} [{...]

varname, *varname_1*, *varname_2*	Variablennamen

Beispiel:

```
missing values  nieder01 to hoehe ozon mess1 to mess4 (9).
```

Hier wird mit dem MISSING VALUES - Kommando für alle aufgelisteten Variablen die 9 als MD-Indikator vereinbart.

17 Literaturverzeichnis

Wallis, W.A. & Roberts, H.V. (1956). Statistics, a new approach. Glencoe, Ill.: The Free Press.

Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation*. Berlin: Springer.

Pedhazur, E.J. & Pedhazur Schmelkin L. (1991). *Measurement, design, and analysis. An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum.

Schnell, R., Hill, P. B. & Esser, E. (1989). *Methoden der empirischen Sozialforschung* (2. Aufl.). München: Oldenbourg.

Kahneman, D. & Miller, D.T. (1986) Norm theory: comparing reality to its alternatives. *Psychological Review*, 93, 136-153.

Scheier, M.F. & Carver, C.S. (1985). Optimism, Coping, Health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4, 219-247.

Bauer, F. (1986) *Datenanalyse mit SPSS* [2. Auflage]. Berlin: Springer.

Bortz, J. (1977). *Lehrbuch der Statistik*. Berlin: Springer.

Hartung, J. (1989). *Statistik* [7. Auflage]. München: Oldenbourg.

Siegel, S. (1976). *Nichtparametrische statistische Methoden*. Frankfurt: Fachbuchhandlung für Psychologie

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. Behavior Research Methods, *Instruments & Computers*, 28, 1-11.

Mehta, C.R., Patel, N.R. (1996). *SPSS Exact Tests 7.0 for Windows*. Chicago, IL: SPSS Inc.

SPSS Inc. (1998). *SPSS Interactive Graphics 8.0*. Chicago, IL.

Stichwortregister